

Aalto-yliopisto
Perustieteiden korkeakoulu
Tietotekniikan tutkinto-ohjelma

Ville Hakola

Videosekvenssin pituuden ja vääristymän sijainnin vaikutus havaittuun laatuun

Diplomityö
Espoo, 21. marraskuuta 2013

Valvoja: Professori Pirkko Oittinen, Aalto-yliopisto
Ohjaaja: TkT Mikko Nuutinen, Aalto-yliopisto

Aalto-yliopisto
 Perustieteiden korkeakoulu
 Tietotekniikan tutkinto-ohjelma

DIPLOMITYÖN
 TIIVISTELMÄ

Tekijä:	Ville Hakola		
Työn nimi:	Videosekvenssin pituuden ja vääristymän sijainnin vaikutus havaittuun laatuun		
Päiväys:	21. marraskuuta 2013	Sivumäärä:	71
Professuuri:	Mediatekniikka	Koodi:	T-75
Valvoja:	Professori Pirkko Oittinen, Aalto-yliopisto		
Ohjaaja:	TkT Mikko Nuutinen, Aalto-yliopisto		
<p>Aikaisemmissa videon laatututkimuksissa on saatu selville, että videosekvenssin loppupään laatu vaikuttaa videon muita osia enemmän kokonaislaatu-arvioon. Nämä tutkimukset ovat kuitenkin koskeneet pääosin vain tietyn pituisia videoita. Työn tavoitteena oli selvittää, miten videosekvenssin pituus vaikuttaa kokonaislaatu-arvioon, kun vääristymä on sijoitettu videosekvenssin alkuun, keskelle tai loppuun.</p> <p>Tutkimusta varten kuvattiin videoita, jotka mahdollistivat eripituisten videosekvenssien vertailun, koska videokontenttien sisältö ei vaihdellut merkittävästi sekvenssin aikana ja koska sekvenssit olivat tarpeeksi pitkiä. Tulosten perusteella vääristyneen osuuden suhteellinen pituus videosekvenssissä vaikutti siihen, miten voimakkaasti sekvenssin loppupäätä painotettiin arvioissa. Videon pituudella (10s, 20s tai 40s) ei tulosten perustella ollut vaikutusta tähän. Jos vääristymän pituus oli puolet videosekvenssin pituudesta, sekvenssin lopussa sijainneella vääristymällä oli alentava vaikutus laatu-arvioihin. Jos vääristymän pituus oli neljäsosa tai kahdeksasosa sekvenssin pituudesta, vääristymän sijainti ei vaikuttanut havaittuun videon laatuun.</p> <p>Tutkimuksen tulos osoittaa, että videon laatua mittaavien algoritmien suorituskykyä voidaan parantaa, jos videokuvia painotetaan eri suhteessa. Tutkimuksen perusteella painottava funktio on laadittava suhteessa vääristymän suhteelliseen pituuteen ja sen sijaintiin. Työssä pohdittiin, että mahdollisesti vääristymän voimakkuus on myös otettava huomioon funktion laadinnassa.</p> <p>Tutkimuksessa luotiin videopankki, joka sisältää tutkimuksessa tuotetut videot sekä subjektiivisen laatu-arvioidatan. Videopankki on vapaasti käytettävissä, ja sitä voi hyödyntää jatkotutkimuksissa.</p>			
Asiasanat:	Videon laatu, videon laadun arviointi, videopankki, videon temporaalinen laatu		
Kieli:	Suomi		

Aalto University
School of Science
Degree Programme of Computer Science and Engineering

ABSTRACT OF
MASTER'S THESIS

Author:	Ville Hakola	
Title:	The impact of video sequence length and distortion position in perceived quality	
Date:	November 21, 2013	Pages: 71
Professorship:	Media Technology	Code: T-75
Supervisor:	Prof. Pirkko Oittinen, Aalto University	
Instructor:	Dr. Mikko Nuutinen, Aalto University	
<p>Previous video quality studies have found that quality at the end of video sequence has higher impact in overall video quality assessment than quality of other parts of sequence. These studies have although only concerned specific video lengths. The aim of this thesis was to study how video sequence length affects in overall video quality assessment when the distorted frames are at the beginning, in the middle or at the end of sequence.</p> <p>The videos, shooted for this study, made it possible to compare different video sequence lengths, because the video content did not change during the sequence and because the sequences were long enough. The results showed that relative distortion length have an effect on how strongly the sequence end is weighted in assessment. The sequence length (10s, 20s or 40s) did not have an effect on this. If the distortion length was half of sequence length, distortions at the end of sequence have degrading effect in assessment. If the distortion length was quarter or eighth of sequence length, the distortion position did not have an effect on perceived video quality.</p> <p>The results show that accuracy of video quality assessment algorithms can be improved if video frames are weighted by weighting function. Based on this research, the function has to be made in relation to relative distortion length and its position. It was discussed that distortion intensity needs possibly also to be taken into account in weighting function formation.</p> <p>The video database, created for this study, consists of produced videos and subjective assessment data. The video database is free to use and it can be utilized for further research.</p>		
Keywords:	Video quality, video quality assessment, video database, temporal video quality	
Language:	Finnish	

Alkusanat

Tämä työ tehtiin Aalto-yliopiston perustieteiden korkeakoulun Mediatekniikan laitoksella.

Kiitän aiheen tarjoamisesta valvojaa Prof. Pirkko Oittista. Hänen järjestämä Viestintätekniikan tutkimusseminaarikurssi oli erityisesti hyödyllinen koko työprosessin kannalta. Samaisen kiitoksen saa myös työn ohjaajaa TkT Mikko Nuutinen, jonka apu oli asiantuntevaa ja nopeasti saatavaa.

Kiitän myös kaikkia kollegoita, kavereita ja perheenjäseniä, jotka ovat tukeneet työn teossa. Kollegoista erityismaininnan saavat Samuli Vuorinen, joka auttoi videoiden kuvaamisessa, sekä Arto Rusanen, jonka kanssa vietetyt kahvihetket saivat ajatukset diplomityön teosta välillä muualle. Kiitos myös kaikille testeihin osallistuneille koehenkilöille.

Espoo, 21. marraskuuta 2013

Ville Hakola

Lyhenteet ja termit

Dekoodaus	Prosessi, joka purkaa koodatun videon esitystä varten
Enkoodaus	Prosessi, joka valmistaa videon sopivaan muotoon tallennusta ja esitystä varten
H.264	H.264/MPEG-4 Part 10 tai AVC (Advanced Video Coding), videonpakkausstandardi
IP	Internet Protocol, protokolla, joka huolehtii tietoliikennepakettien toimittamisesta Internetissä
ISO-arvo	Kuvakennon herkkyyttä valolle kuvaava lukema
Laatuattribuutti	Laatuun vaikuttava tekijä
MPEG-2, MPEG-4	Motion Picture Experts Groupin kehittämät videonpakkausstandardit
Referenssivideo	Alkuperäinen vääristymätön video
Temporaalinen	Aikaan liittyvä, ajallinen
Videokontentti	Kuvattu videosisältö
Videosekvenssi	Alkuperäisestä videosta leikattu tietyn pituinen video
Videon havaittu laatu	Havaittajan mielipide videon laadusta
Videon tuotettu laatu	Videosysteemin tuottaman videon laatu
Videon laatu	Systeemin läpi menneen videon ominaispiirre, joka mittaa havaitun videon vääristyneisyyttä
Videon versio	Videosekvenssistä luotu vääristynyt video

Sisältö

Tiivistelmä	2
Abstract	3
Alkusanat	4
Lyhenteet ja termit	5
1 Johdanto	8
1.1 Taustaa	8
1.2 Työn tavoite ja tutkimuskysymykset	9
1.3 Työn rakenne ja tutkimusprosessi	10
2 Videon laatu	11
2.1 Kuvanlaadusta videon laatuun	11
2.2 Videon laadun muodostuminen	13
2.2.1 Kuvaus	14
2.2.2 Pakkaus	16
2.2.3 Siirto	17
2.2.4 Näyttäminen	18
2.2.5 Havainnointi	18
3 Videon laadun arviointi	21
3.1 Subjekttiivinen videon laadun arviointi	21
3.2 Objektiiviset mitat	22
3.2.1 Objektiivisten mittojen luokittelu	22
3.2.2 Täyden referenssin videon laatumittojen laskeminen	25
3.3 Temporaalinen videon laadun arviointi	26
4 Videopankit	30
4.1 Videopankkien piirteet	30

4.2	Mitta-arvot videoiden karakterisointiin	32
5	Tutkimusmenetelmät	35
5.1	Testivideot	35
5.2	Testimenetelmä	40
5.3	Materiaalin valinta	43
5.3.1	Esitesti I	43
5.3.2	Esitesti II	45
5.4	Testi I: vakiopituinen vääristymä	48
5.5	Testi II: suhteellinen vääristymä	50
6	Tulokset ja analyysi	52
6.1	Testi I: vakiopituinen vääristymä	52
6.2	Testi II: suhteellinen vääristymä	55
7	Johtopäätökset	59
	Lähdeluettelo	60
	Liite A Esitestit	68
	Liite B Testit I ja II	69

Luku 1

Johdanto

1.1 Taustaa

Video on sarja kuvia, jotka ihminen voi havaita näyttölaitteelta. Videoita katsotaan maailmanlaajuisesti keskimäärin yli kolme tuntia vuorokaudessa ja katseluaikat ovat vain nousussa [7]. Reilusti eniten niitä katsotaan edelleen perinteisestä televisiosta [39]. Voimakkaimmassa nousussa ovat kuitenkin tietokoneilta, tabletti- ja mobiililaitteilta katsottujen videoiden osuudet [2].

Videon laadun tutkimus eroaa still-kuvien tutkimuksesta siinä mielessä, että videoissa on temporaalinen eli ajallinen ulottuvuus. Tällöin ihminen jo lähtökohtaisesti prosessoi informaatiota eri tavalla kuin staattisissa kuvissa. Videoissa saattaa myös esiintyä sellaisia vääristymiä, joita still-kuvissa ei ole. Merkittävää on myös se, että videon laatu saattaa vaihdella videon katselun aikana.

Dynaaminen streemaus on tekniikka [32], jossa videon laatu vaihtelee videosekvenssin aikana. Siinä loppukäyttäjän vastaanottaman videon laatua muunnetaan dynaamisesti muun muassa saatavilla olevan kaistanleveyden ja paikallisen laitteiston tehokkuuden mukaan. Tekniikka edellyttää, että videoista on muodostettu useita erilaatuisia versioita, joissa muun muassa resoluutio tai pakkaustehokkuus voi vaihdella. Dynaaminen streemaus mahdollistaa perinteistä toistoa lyhyemmät puskurointiajat, nopeammat toiston aloitukset sekä sujuvan katselun eritasoisilla laitteilla.

Videon laatu voi vaihdella sekvenssin aikana myös, jos videota kuvatesa, pakatessa tai siirrettäessä siihen muodostuu vääristymiä. Niitä havaitaan muun muassa videon sumentumisena, virheellisinä väreinä ja toiston nykimisenä. Vääristymät voivat vaikuttaa videonlaatuun niin spatiaalisesti kuin temporaalisesti. Vääristymien tyyppi, kesto, voimakkuus, sijainti ja yleisyys muodostavat lopullisen videon laadun.

Videon laatua voi mitata subjektiivisin tai objektiivisin menetelmin. Luotettavimman laatuarvion mahdollistavat subjektiiviset menetelmät, koska niissä arviot vastaavat sitä, mitä koehenkilöt oikeasti havaitsevat. Objektiiviset eli algoritmeihin pohjautuvat menetelmät ovat yleensä helpommin, halvemmin ja nopeammin toteutettavissa, mutta niillä ei välttämättä saavuteta samaa tarkkuutta arviossa kuin subjektiivisissa menetelmissä.

Videon laadun arviointia varten on luotu videopankkeja. Ne ovat videokoelmia, jotka tyypillisesti sisältävät useita videokontentteja, joista on luotu erilaisia vääristyneitä versioita. Usein myös versiokohtaiset laatuarviot ovat saatavilla. Videopankkien avulla objektiivisia algoritmeja pystytään kehittämään ja testaamaan.

Ihmisellä on tapana unohtaa kokemuksia, jotka tapahtuivat kauan aikaa sitten, ja painottaa havaintoja, jotka tapahtuivat äskettäin [42]. Tämä muistiefekti on otettava huomioon videon laatua arvioivien algoritmien laadinnassa. Yksittäisien videokuvien laaduista laskettu keskiarvo ei aina ennusta tarkasti videon kokonaislaatua.

Subjektiivisista testeistä [4, 15, 35, 68] saatujen tuloksien perusteella videon lopussa sijaitseva vääristymä koetaan häiritsevämpänä kuin vastaavan pituinen vääristymä videon alussa. Kaukana videon lopusta sijaitsevien vääristymien sijainti ei enää vaikuta kokonaislaatuarvioihin.

1.2 Työn tavoite ja tutkimuskysymykset

Edelliset tutkimukset [4, 15, 35, 68] ovat mitanneet havaittua videon laatua, kun videosekvenssissä olleen vääristymän tai heikomman laadun sijainti on vaihdellut. Sen sijaan itse videosekvenssin pituuden vaikutusta tilanteeseen ei ole kattavasti tutkittu. Työn tavoitteena on selvittää näiden parametrien yhteisvaikutusta.

Työn tutkimuskysymykset voidaan esittää seuraavassa muodossa:

- Miten videosekvenssin pituus ja vääristymän sijainti vaikuttavat havaittuun laatuun?
- Vastaavatko tulokset kirjallisuudessa esitettyjä havaintoja temporaalisesta videon laadusta?

Työn tutkimus ei kohdistu suoranaisesti mihinkään tiettyyn videopalveluun, vaan tuloksia on mahdollista soveltaa uusien videopalveluiden suunnittelussa ja jo olemassa olevien palveluiden laadun kehittämisessä. Tutkimuksen tulokset voivat toimia pohjana myös videon laatua arvioivien algoritmien laadinnassa.

Tulokset ovat hyvin sovellettavissa palveluihin, joissa videon laatu vaihtelee ajan myötä. Videopalveluista esimerkiksi Youtube¹ ja Netflix² käyttävät hyödyksi dynaamista streemaustekniikkaa, jossa video saattaa alkaa toistua heikolla laadulla, mutta muuttuu ajan myötä paremmaksi, kun korkealaatuinen video on latautunut puskuriin.

1.3 Työn rakenne ja tutkimusprosessi

Tämä työ on jaettu kirjallisuusosaan (Luvut 2–4) ja kokeelliseen osaan (Luvut 5–6). Luvussa 2 esitetään videon laatuun vaikuttavia tekijöitä, ja luvussa 3 videon laadun arviointimenetelmiä. Omassa alaluvussaan (3.3) tarkastellaan temporaalisen videon laadun arviointia. Luvussa 4 kerrotaan videopankkeista.

Kokeellinen osa alkaa luvusta 5. Siinä kerrotaan tutkimusta varten luodusta videopankista sekä testimenetelmästä. Tulokset ja analyysi on esitetty luvussa 6. Johtopäätökset on esitetty luvussa 7.

¹<http://www.youtube.com/>

²<https://www.netflix.com/>

Luku 2

Videon laatu

Tässä luvussa esitetään piirteitä, jotka vaikuttavat videon laatuun.

2.1 Kuvanlaadusta videon laatuun

Sanakirjan määritelmän¹ mukaan kuvanlaatu on kuvan ominaispiirre, joka mittaa havaitun kuvan vääristyneisyyttä. Tyypillisesti kuvantamissysteemit aiheuttavat kuvaan vääristymiä.

Keelanin [28] mukaan kuvanlaatu on havainnoitsijan havaitsema kuvan arvo tai sen erinomaisuus. Keelanin mukaan kuvan arvo muodostuu kuvanlaatuattribuuteista, jotka voivat olla joko artefakteihin, mieltymyksiin tai esteettisyyteen liittyviä.

Artefakteihin liittyviä attribuutteja syntyy kuvauksessa, pakkauksessa ja kuvaa siirrettäessä ja näitä ovat esimerkiksi kuvan epäterävyys ja rakeisuus. Mieltymyksiin liittyvät attribuutit ovat yleensä riippuvaisia sekä havainnoitsijan mielipiteistä että kuvan sisällöstä. Näitä ovat esimerkiksi kuvan kontrasti ja värikkyys. Esteettisyyteen liittyvät attribuutit ovat edellä mainittuja attribuutteja vielä riippuvaisempia subjektiivisesta havainnosta, joten niiden mittaaminen on yhä vaikeampaa. Esteettisyyttä kuvaavia attribuutteja ovat esimerkiksi kuvan sommittelu ja kameran suunta. [28]

Keelan ei sisällytä kuvanlaadun määritelmään henkilökohtaisiin piirteisiin liittyviä attribuutteja, koska nämä attribuutit eivät ole helposti käsiteltävissä tai mitattavissa ja koska ne eivät ole tärkeitä kuvantamissysteemin suunnittelun kannalta. Kyseisiä attribuutteja ovat muun muassa niitit, jotka herättävät havainnoitsijassa muistoja tai vaikuttavat tunteisiin.

Kuvan laatuattribuutit ovat piirteitä, joilla on vaikutusta kuvanlaadussa. Kuvan laatuattribuutit voidaan jakaa alhaisen ja korkean tason laatuatt-

¹<http://encyclopedia.thefreedictionary.com/image+quality>

ribuutteihin [40]. Alhaisen tason laatuattribuutit ovat konkreettisempia ja helpommin arvioitavissa kuin korkean tason attribuutit. Korkean tason attribuutit ovat puolestaan enemmän riippuvaisia subjektiivisesta mielipiteestä. Laatuhierarkioiden [9, 18] avulla pyritään kuvaamaan tasot ja yhteydet laatuattribuuttien välillä. Korkeatasoisin attribuutti on tyypillisesti kokonaislaatu, joka jakaantuu alemman tason attribuutteihin. Laatuattribuuteilla voidaan arvioida sitä, miten muutokset alhaisen tason attribuuteissa vaikuttavat ylemmällä tasolla tai sitä, miten ylempään tason laatuattribuutit muodostuvat alemman tason attribuuteista [13].

Radun ym. [46] tutkivat kameran kuvanprosessointiketjun vaikutusta havaitussa laadussa. Heidän mukaansa korkealaatuisen kuvan korkeantason vaatimuksia ovat luonnollisuus ja selkeys. Luonnollisuus liittyy kuvan ja havainnoitsijan odottaman tai todellisen näkymän suhteeseen. Selkeydellä puolestaan tarkoitetaan tässä sitä, kuinka helppoa havainnoitsijan on erottaa kuvan sisältö. Radunin ym. mukaan kuvanlaatu heikentyy eri syistä, jos kuva ei ole korkealaatuinen. Tällöin kuvan häiriöt liittyvät alhaisen tason laatuattribuutteihin, kuten tummuus ja epäterävyys.

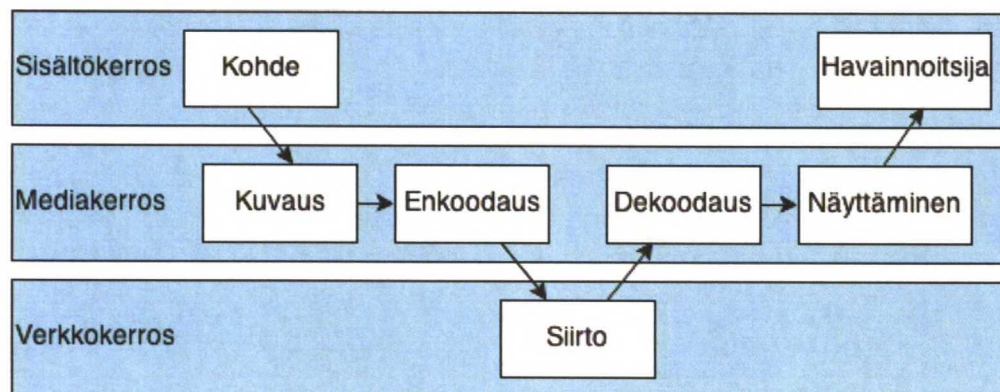
Nymanin ym. [41] tutkimuksessa koehenkilöt kuvailivat kameroilla otettujen videoiden ominaisuuksia. Yleisimmät alhaisen tason laatuattribuutit liittyivät terävyyteen, kohinaan, valaistukseen ja väreihin. Murdochin ym. [38] tutkimuksessa arvioitiin puolestaan televisiokuvan laatua, kun katseluetäisyyttä ja ympäröivää valaistusta vaihdettiin. Tutkimuksen mukaan kirkkaus, kontrasti, väri, terävyys ja liikkeen laatu olivat tärkeimmät laatuattribuutit televisiokuvassa.

Kuvanlaadun tekijät on yhdistettävissä videon laatuun, koska video on sarja kuvia. Formaali videon laadun määritelmä onkin kuvien määritelmään nähden hyvin samankaltainen: *Videon laatu on systeemin läpi menneen videon ominaispiirre, joka mittaa havaitun videon vääristyneisyyttä*.² Videon laadusta puhuttaessa on kuitenkin otettava huomioon myös tempo-
raalinen eli ajallinen ulottuvuus, joten video havaitaan jo lähtökotaisesti eri tavalla kuin kuvat. Videoissa saattaa esiintyä sellaisia vääristymiä, joita kuvissa ei ole. Lisäksi videon laatu saattaa muuttua sekvenssin aikana merkittävästi. Tietyllä ajanhetkellä videon laatu voi olla vielä hyväksyttävää, mutta toisaalla muuttua täysin katselukelvottomaksi.

²<http://encyclopedia.thefreedictionary.com/video+quality>

2.2 Videon laadun muodostuminen

Wikstrand [64] jakaa videon laatuun vaikuttavat tekijät kolmelle abstraktille tasolle: sisältö-, media-, ja verkkokerrokselle. Kuvassa 2.1 on nähtävillä nämä kerrokset ja ketjurakenne siitä, miten videon laatu muodostuu tyypillisessä sovelluksessa, jossa video kuvataan enkoodataan, siirretään, dekoodataan ja näytetään. Rakenteen eri komponenteissa syntyy erilaisia vääristymiä videoon.



Kuva 2.1: Videon laadun muodostuminen abstrakteilla kerroksilla [64].

Sisältökerroksella käsitellään tyypillisesti videon sisällön vaikutusta vastaanottajassa. Mediakerroksella tapahtuu lähettäjän päässä videon kuvaus sekä enkoodaus ja vastaavasti vastaanottajan päässä videon dekoodaaminen sekä näyttäminen. Verkkokerros keskittyy videon siirtoon. Rakenteesta on esimerkkinä muun muassa digitaaliset televisiolähetykset. Aluksi sisältö tai kohde kuvataan videokameralla, jonka jälkeen video käsitellään ja pakataan lähetettäväksi siirtoverkkoon. Vastaanottajan päässä verkosta tullut videovirta puretaan ja esitetään käyttäjälle esimerkiksi televisiolta, tietokoneelta, tabletista tai mobiililaitteesta.

Käyttäjakeskeisen ajattelutavan mukaan videon laatu nähdään yhdistelmänä videon tuotetusta laadusta (engl. produced quality) ja videon havaitusta laadusta (engl. perceived quality). Videon tuotettu laatu muodostuu videosisältöön ja -systeemiin liittyvistä tekijöistä, kun puolestaan videon havaittu laatu on ihmisen havainnollisten prosessien tuotos [26, 27]. Wikstrandin mallissa [64] videon tuotettu laatu muodostuu media- ja verkkokerroksilla ja videon havaittu laatu ylimmällä sisällökerroksella. Havaittuun laatuun vaikuttaa tietenkin myös se, mitä alimmilla kerroksilla tapahtuu. Mallin eri

kerroksien rajoja ei voi kuitenkaan nähdä kovin tiukkoina.

Havaitun ja tuotetun videon laadun kanssa samankaltaisia termejä ovat kokemuksen laatu (engl. quality of experience) ja palvelun laatu (engl. quality of service). Kokemuksen laatu on käyttäjäkeskeinen termi ja se mittaa sovelluksen tai palvelun yleistä hyväksyttävyyttä subjektiivisesti eli niin miten loppukäyttäjä on sen havainnut [24]. Kokemuksen laatuun vaikuttaa systeemin lisäksi myös käyttäjän odotukset ja konteksti. Palvelun laatu puolestaan pitää sisällään kaikki palvelun piirteet, jotka vaikuttavat sen kykyyn tyydyttää käyttäjän tarpeet [22]. Tyypillisesti videon laatu on vain osa videopalvelun laatua, jolloin onkin yleensä perusteltua tutkia koko palvelun laatua vuorovaikutustilanteineen.

Videon tuotettuun laatuun vaikuttavat komponentit ovat videon kuvaus, pakkaus ja näyttäminen. Videon pakkauksen katsotaan sisältävän enkoodauksen, jossa video koodataan; ja dekodauksen, jossa videopakkaus puretaan. Tässä luvussa painotutaan erityisesti siihen, minkälaisia vääristymiä prosessin eri vaiheissa saattaa muodostua, jotka vaikuttavat videon laatuun.

2.2.1 Kuvaus

Tässä kuvaus sisältää optiikan, sensorin ja kuvanprosessoinnin. Optiikka muodostaa näkymän sensorille, joka muuttaa signaalin sähköiseksi ja edelleen käsiteltäväksi. Kuvanprosessointi muokkaa signaalin muotoon, joka voidaan tallentaa ja esittää näyttölaitteella.

Optisia vääristymiä eli linssivirheitä voi syntyä kamerassa, kun valo kulkeutuu linssien läpi. Optiset vääristymät voidaan jakaa geometrisiin ja kromaattisiin linssivirheisiin. Geometriset linssivirheet liittyvät linssin pallomaiseen muotoon, kun puolestaan kromaattiset linssivirheet liittyvät valon lukuksiin eri taajuuksiin. [43, 60]

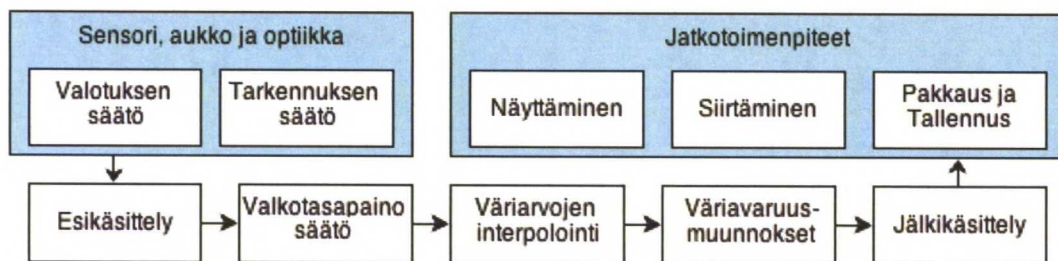
Geometrisiä linssivirheitä ovat muun muassa palloaberraatio, koma, astigmatismi, kuvapinnan kaarevuus ja geometriset vääristymät. Palloaberraatiossa linssin reunalle tulevat valonsäteet taittuvat eri etäisyydelle linssistä kuin linssin keskiosaan tulevat säteet. Tämä ilmenee kuvan sumentuneisuutena, koska säteillä ei ole yhteistä polttopistettä optisella akselilla. Koma on hyvin palloaberraation kaltainen linssivirhe. Siinä linssin keskipisteen ulkopuolelta kuvautuvan objektin reunasäteet muodostavat kuvan eri kohtaan kuvatasolle kuin linssin keskeltä kulkevat säteet. Koman voi nimensä mukaisesti havaita objektien komeettamaisena pyrstönä. [43]

Astigmatismissa puolestaan valonsäteet taittuvat eri etäisyydelle linssistä riippuen siitä, tulevatko ne sinne vaaka- vai pystytason suuntaisina. Astigmatismien voi havaita pisteiden epäterävyytenä. Kuvapinnan kaarevuudella tarkoitetaan sitä, että kuvapinta on todellisuudessa kulhon muotoinen

eikä kuvaa pystytä muodostamaan koko alueeltaan tarkasti tasomaiselle pinnalle. Tällöin käytetään usein termejä tyyny- ja tynnyrivääristymä. Näissä vääristymätyypeissä kuva on tarkka, mutta geometrisesti vääristynyt, jolloin suorat viivat kaareutuvat joko sisään- tai ulospäin. [60]

Kromaattiset, eli väreihin liittyvät linssivirheet voidaan jakaa aksiaalsiin ja poikittaisiin värivirheisiin. Aksiaalisessa värivirheessä objektin eri värit taittuvat eri kohdille optista akselia. Poikittaisessa värivirheessä kuva puolestaan muodostuu yhtenäiselle tasolle, mutta eri etäisyydelle optisesta akselistä. Kromaattiset virheet aiheuttavat värien sumentuneisuutta ja yksittäisten värirantujen muodostumisia kuvassa. Ne johtuvat siitä, että linssi taittaa valoa eri tavalla riippuen siihen kohdistuneesta valon aallonpituudesta. [60]

Valotuksen säädössä kamera muodostaa optimaalisen aukkokoon, valotusajan ja ISO-arvon niin, että kuvan kirkkaat kohteet eivät saturoidu ja että tummien kohteiden yksityiskohdat erottuvat. Tarkennuksen säädössä linssi-systeemin paikka säädetään niin, että kohde tarkentuu sensorille. Tämän jälkeen kuvassa 2.2 esitetty kameran kuvanprosessointiketju ottaa syötteenä sensorin muodostaman raakadatan ja luo siitä digitaalisen kuvan, joka voidaan pakata sekä tallentaa pysyväan muistiin ja näyttää käyttäjälle. [3]



Kuva 2.2: Kuvanprosessointiketju [47].

Esikäsittelyvaiheessa sensorista saadulle raakakuvalle säädetään mustan taso. Tämä tapahtuu tyypillisesti vähentämällä täysin musta kuva sensorin raakadatasta. Epäonnistunut mustan tason säätö aiheuttaa kontrastihäviötä kuvassa. Optiset, elektroniset, digitaaliset ja sähköiset lähteet saatavat tuottaa kohinaa kuvaan, jolloin myös kohinansuodatus on tarpeellista esikäsittelyvaiheessa. Liian voimakas suodatus voi kohinan vähentämisen lisäksi sumentaa kuvaa. [3, 47]

Valkotasapainoa säätämällä eri valaistusolosuhteissa otetut kuvat pyritään saada vastaamaan ihmissilmän havaintoa. Esimerkiksi valkoinen paperi koetaan valkoisena, vaikka se olisi punertavassa valaistuksessa. Tähän kamera ei adaptoidu samalla tavalla kuin ihmissilmä, koska eri valaistuksis-

sa objektien radianssi on erilainen. Epäonnistuneen valkotasapainon säädön tuloksena kuvan valkoinen ei ole valkoista. [47]

Väriarvojen interpoloinnissa muodostetaan suuruudet puuttuville väriarvoille. Väriarvoja puuttuu, koska yleensä jokainen pikseli sensorilla edustaa vain yhtä kolmesta värikomponentista. Interpolointitekniikat käyttävät hyödyksi naapuripikseleiden informaatiota puuttuvien väriarvojen ennustuksessa. [3, 47]

Väriavaruusmuunnosvaiheessa värit joudutaan aluksi muuntamaan sensorivaruudesta laiteriippumattomaan avaruuteen, kuten CIEXYZ-väriavaruuteen, koska eri sensorit tuottavat erisuuruisia väriarvoja. Tästä avaruudesta värit tyypillisesti muunnetaan edelleen laiteriippuvaan väriavaruuteen, kuten sRGB-väriavaruuteen. Tämä tehdään, jotta kuvat havaitaan samanlaisina samaa standardia noudattavilla näytöillä. Väriavaruusmuunnosvaiheissa eritasoisten väriväristymien syntyminen on mahdollisia. [47]

Jälkikäsittelyvaiheessa voidaan vielä suorittaa erilaisia toimenpiteitä, joilla pyritään parantamaan kuvanlaatua. Esimerkiksi kuvien reunoja terävöitetään usein, koska ihmissilmä on kaikista adaptiivisin teräville reunoille. Toinen yleinen toimenpide on kontrastin parannus, jossa pyritään löytämään optimaalisin kontrasti ja kirkkaus kuvalle. [3]

2.2.2 Pakkaus

Videodatan säilyttäminen vaatii paljon kovalevytilaa ja siirtäminen paljon kaistaa. Pienentämällä videotiedoston kokoa häviämättömien pakkausalgoritmien avulla saavutetaan yleensä vain puolet pienempi tiedostokoko, mikä ei ole riittävästi. Käytännössä video pakataan aina häviöllisiä menetelmiä käyttäen. Näissä menetelmissä hyödynnetään usein sitä, että videon pikselit korreloivat naapuripikselien kanssa niin spatiaalisesti kuin temporaalisesti ja että havainnoitsijalle ei kaikki videoinformaatio ole näkyvää. [65]

Tyypillinen videonpakkausalgoritmi sisältää muunnos-, kvantisointi- ja koodausvaiheen. Muunnosvaiheessa videokuvat jaetaan pienempiin alueisiin eli blokkeihin ja ne muunnetaan taajuusavaruuteen. Tässä avaruudessa voidaan erottaa eri taajuuskaistat, joille ihmissilmä on herkkä. Kvantisointivaiheessa taajuuksien esittämistarkkuutta pienennetään. Tyypillisesti ihmiselle vaikeasti havainnoitavia korkeataajuisia komponentteja kvantisoidaan enemmän kuin alhaisien taajuuksien komponentteja. Koodausvaiheessa data voidaan pakata häviöttömästi hyödyntäen kvantisoitujen kertoimien redundanssia. [65]

Minimoidakseen redundanssia yksinkertaisessa videon pakkausmenetelmässä koodataan vain kuvien väliset pikselikohtaiset eroavaisuudet. Parempaan pakkausasteeseen päästään kuitenkin käyttämällä menetelmiä, joilla

pyritään ennustamaan objektien liikettä esimerkiksi liikevektoreiden avulla. Tätä hyödynnetään useissa videon pakkausstandardeissa, kuten yleisessä H.264:ssä. [48]

Monissa videon pakkausmenetelmissä käytetään niin sanottuja I-, P- ja B-kuvia. I-kuvat ovat referenssikuvia ja ne sisältävät itsessään kaiken kuvan muodostukseen tarvittavan informaation. P-kuvat käyttävät sen sijaan videosekvenssissä edeltävien I- tai P-kuvien informaatiota hyödyksi kuvan muodostuksessa. B-kuvat käyttävät edellisten kuvien informaation lisäksi ennustuksen apuna myös tulevien kuvien informaatiota. Koska I-kuvat ovat täysin valmiita kokonaisina kuvia, ne ovat pakkaustehokkuudeltaan kaikista heikoimmat. B-kuvat saadaan usein pakattua kaikista pienimpään tilaan, mutta niissä on itsessään kaikista vähiten kuvainformaatiota. Videosekvenssissä GOP (Group of Pictures) eli kuvien rakenne, joka toistuu voi esimerkiksi olla IBBPBBPBBPBB. [48]

Korkea kvantisointiaste saattaa aiheuttaa blokkisuutta videossa, jonka voi havaita irrallisina blokin kokoisina laatikkoina kuvassa tai reunakohtien jatkumattomuudella. Sumentuneisuutta tai niin sanottuja Gibbsin artefakteja voi havaita, jos videosta on poistettu paljon korkeataajuisia komponentteja. Gibbsin artefakti ilmenee muotojen tai etenkin terävien reunojen toistumisena niiden ympärillä. [65]

Temporaalisia artefakteja ovat muun muassa hyttyskohina ja nykivä liike. Hyttyskohina muistuttaa hyttysiä, jotka parveilevat videon yllä. Sitä ilmenee, kun tasainen tekstuuri koodataan eri tavalla peräkkäisissä kuvissa. Nykivää liikettä saattaa puolestaan esiintyä, kun pakkausalgoritmin liikkeentunnistus ei toimi täysin optimaalisella tavalla. [65]

2.2.3 Siirto

Verkkokerroksella verkon vajaatoiminta aiheuttaa erilaisia artefakteja videoon. Tyypillinen verkon piirre on se, että videopaketit eivät siirry siellä tasaisin väliajoin. Tätä epätasaisuutta pyritään välttämään puskurien käytöllä, mutta ne aiheuttavat viivettä pakettien siirrosta. Kaiken lisäksi jos puskuri täyttyy, paketti voi kadota kokonaan [36]. Siirtokerroksella voi ilmetä myös bittivirheitä, joita tyypillisesti esiintyy, kun videota siirretään langattomasti. Näitä virheitä saattavat aiheuttaa muun muassa vaimea signaali tai muiden laitteiden häiritsevät signaalit [56].

Viivästyneet paketit havaitaan viiveenä videon toistossa, ja kadonneet paketit havaitaan useina erilaisina artefakteina [30]. Näitä ovat esimerkiksi väärin sijoittuneet objektien reunat, virheelliset värit sekä sumentuneet kuvan alueet.

Koska videon pakkaamisessa käytetään usein P- ja B-kuvia, paketin puutoaminen ei yleensä aiheuta vääristymiä vain paketin omassa lohkoissa, vaan myös kaikissa muissa lohkoissa, jotka ovat riippuvaisia tästä lohkoista. Yhden paketin katoamisella saattaa olla laaja vaikutus videon laadussa niin spataalisesti kuin temporaalisesti. [65]

2.2.4 Näyttäminen

Tänä päivänä on olemassa laaja valikoima erikokoisia ja eri tekniikoilla toimivia näyttöjä. Perinteisten kuvaputki-, LCD- ja plasmatekniikkojen lisäksi videoita voi esittää muun muassa projektoreilla. Yleisten videolaitteiden, kuten tietokoneen ja television, kilpailijoiksi ovat nousemassa tabletit ja mobiililaitteet [2]. Laitteiden ominaispiirteet, kuten kuvan kirkkaus, kontrasti ja vasteaika, määräävät lopullisen videosysteemin tuotetun laadun. Jotkut artefaktit saattavat olla havaittavampia tietyn tyyppisiä näyttöjä käytettäessä [65].

Murdochin ym. [38] tutkimuksessa näytön pikselien määrällä on todettu olevan vaikutus havaitussa kuvanlaadussa. Tutkimuksessa on muodostettu funktio, jonka avulla pystytään laskemaan näytön pikselimäärälle optimaalinen näytön koko tietyllä katseluetäisyydellä tai optimaalinen katseluetäisyys tietylle näytön koolle.

Barten ym. [6] tutkimuksessa katseluetäisyydellä ja ympäristön valaistuksella on todettu olevan vaikutus havaitussa videon laadussa. Tutkimuksesta on saatu tuloksena, että videon terävyyden, kirkkauden ja värien vaikutus riippuu katseluetäisyydestä. Ympäristön valaistus on puolestaan vaikuttanut havaittuun videon mustan tasoon, jolloin videon kontrasti on muuttunut.

2.2.5 Havainnointi

Ihmisen silmä on näköelin, joka vastaanottaa valoa ja muuntaa sen hermoimpulsseiksi, jotka edelleen kulkeutuvat näkörataa pitkin aivoihin, jossa näköaistimus syntyy. Valoa aistivia fotoreseptoreja ovat sauva- ja tappisolut, jotka sijaitsevat verkkokalvolla silmän takana. Ihmisellä on kolmenlaisia tappisoluja, jotka aktivoituvat joko punaisesta, vihreästä tai sinisestä valosta. Sauvasolut eivät pysty erottelemaan värejä, mutta ne mahdollistavat hämärässä näkemisen. Verkkokalvolla gangliosolut muokkaavat sauva- ja tappisoluista tullutta informaatiota ja lähettävät sen edelleen hermoimpulsseina näköhermoa ja näköjuostetta pitkin keskellä aivoja sijaitsevaan ulompaan polvitumakkeeseen, jossa näköinformaatiota prosessoidaan. Ulommasta polvitumakkeesta informaatio siirtyy edelleen aivojen takaosaan

näköaivokuorelle, joka on ihmisen näköjärjestelmän suurin osa. Siellä muodostuu lopullinen korkean tason näköaistimus. [12, 52]

Ihmisen näköjärjestelmän vaste riippuu kohteen luminanssista. Kohteen luminanssin suhteella ympäristön luminanssiin on todettu olevan suurempi vaikutus kuin kohteen absoluuttisella luminanssilla. Kohteen havaitsemiseen vaikuttaa suuresti myös kohteen ominaispiirteet, kuten sen väri sekä spatiaalinen ja temporaalinen taajuus. Lisäksi esimerkiksi maskaus ja adaptoituminen ovat tunnettuja ilmiöitä liittyen kohteen havaitsemiseen. Maskaus liittyy siihen, että joitain ärsykettä ei pysty havaitsemaan toisen vielä voimakkaamman ärsykkeen takia. Adaptoituminen on sitä, että ihmisen näköjärjestelmä voi tottua tiettyihin olosuhteisiin, jolloin vasteen herkkyys muuttuu. [65]

Visuaalinen informaatio prosessoidaan eri reittejä pitkin riippuen sen piirteistä, kuten väreistä, spatiaalisesta taajuudesta tai liikkeen suunnasta. Nämä reitit ovat merkittävässä roolissa, kun yritetään tutkia piirteiden yhteisvaikutusta. Vaikka ihmisen visuaalinen systeemi on hyvin adaptiivinen, se ei ole yhtä herkkä kaikille ärsykeille. Kaiken kaikkiaan ihmisen näköjärjestelmä on erittäin monimutkainen systeemi, jota nykyinen tieto ei pysty täysin selittämään. [65]

Videon sisältökerroksella tapahtuvia asioita on erittäin vaikea mitata, koska videon laatu riippuu siellä havainnoitsijasta ja videosisällöstä. Eri-ikäiset, eri sukupuolta tai eri kulttuuria edustavat ihmiset saattavat arvioida saman sisällön täysin eri tavalla. Vaihtelevuutta arviointeihin tuovat havainnoitsijoiden sensoriset ominaisuudet, henkilökohtaiset odotukset, kokemus ja motivaatio. Arviointikriteerit muuttuvat esimerkiksi sen mukaan, arvioidaanko TV-mainoksia vai urheiluohjelmia. [34] Kuvan havaittuun laatuun vaikuttaa myös se, herättääkö kuva miellyttäviä muistoja tai tunteita havainnoitsijassa [28].

Multimodaalinen havaitseminen on prosessi, missä kahdesta tai useammasta aistinelimestä tuleva informaatio yhdistetään havaintokokemukseksi aistijärjestelmässä [29]. Tämä on monimutkainen prosessi, missä eri lähteiden informaatio täydentää ja muokkaa lopullista kokemusta. Multimediakontekstissa multimodaalisuus liittyy usein videon ja äänen vuorovaikutukseen.

Videon laadun havaitsemiseen tuovat poikkeavuutta erilaiset kognitiiviset vääristymät. Ne liittyvät ihmisen taipumukseen painottaa havaintoja tai informaatiota tavoilla, jotka johtavat epäjohdonmukaisiin arvioihin tai tulkintoihin. Videon temporaaliseen havaitsemiseen liittyviä kognitiivisia vääristymiä ovat muun muassa muistiefekti, pituuden laiminlyönti, smoothing-efekti ja epäsymmetrinen vaste.

Tässä työssä **muistiefektiksi** kutsutusta kognitiivisesta vääristymästä on kirjallisuudessa käytetty termejä "forgiveness effect" ja "recency effect". Forgiveness-efekti liittyy siihen, että ihmisellä on tapana unohtaa kokemuk-

sia, jotka tapahtuivat kauan aikaa sitten. Recency-efekti puolestaan viittaa siihen, että ihminen painottaa enemmän havaintoja, jotka tapahtuivat äskettäin. Nämä efektit johtuvat ihmisen lyhytkestoisesta muistista. [42]

Pituuden laiminlyönnillä tarkoitetaan tässä ilmiötä, jossa videon tai vääristymän pituus ei vaikuta havaintoon. Sen sijaan ärsykkeen voimakkuus määrittää ihmisen vasteen suuruuden. [42]

Voimakkaan ärsykkeen vaikutus kestää lyhyen ajan myös ärsykkeen jälkeen vaimentuen kuitenkin koko ajan hiljalleen. Tämä **smoothing-efekti** johtuu ihmisen lyhytkestoisesta muistista. Videosekvenssissä esimerkiksi muutama erittäin vääristynyt kuva vaikuttaa niiden esiintymisen jälkeenkin. Toisin sanoen vääristymä vaikuttaa havaintoon sen esiintymän jälkeenkin, vaikka video olisikin jo visuaalisesti täysin vääristymätöntä. Jos videossa on kaksi vääristynyttä kohtaa lähellä toisiaan, ei ensimmäisen vääristymän jälkeen ihmisen vaste välttämättä ennäätä palata vääristymien välissä olevaa laatua vastaavalle tasolle. Tällöin vääristymien väliset kuvat voidaan arvioida myös vääristyneiksi. [11]

Epäsymmetrinen vaste tarkoittaa sitä, että ihmiset muistavat paremmin epämiellyttäviä kokemuksia kuin miellyttäviä. Tämä havainto on soveltuvin myös videon laadun muutoksiin. Ihmiset kokevat voimakkaammin videon laadun huononemisen kuin objektiivisesti vastaavan videon laadun parantumisen. [11]

Luku 3

Videon laadun arviointi

Videon laatua voi mitata subjektiivisesti tai objektiivisesti. Luvussa 3.1 esitetyissä subjektiivisissa menetelmissä videonlaatu-arvio perustuu havainnoitsijoiden yksilöllisiin näkemyksiin, eli koehenkilöiden mielipiteisiin, joista voidaan muodostaa keskimääräinen laatu-arvio. Luvussa 3.2 esitettyt objektiiviset eli laskennalliset menetelmät arvioivat puolestaan algoritmien avulla videon laatua. Videon temporaalisen laadun arviointiin on keskitytty luvussa 3.3.

3.1 Subjektiivinen videon laadun arviointi

Subjektiivisessa videon laadun arvioinnissa mitataan koehenkilöiden subjektiivista havaintoa videosekvensseistä. Subjektiivinen testaus on tyypillisesti enemmän aikaa vievää ja kalliimpaa järjestää kuin objektiiviset mittaukset, mutta niistä saadut laatu-arviot vastaavat todellista ihmisen havaintoa [21, 25].

Tyypillinen subjektiivisen testaus koostuu ärsykkeiden valinnasta, ärsykkeiden valmistelusta, koehenkilöiden valinnasta, tutkimusmenetelmän valinnasta, testin suorittamisesta ja tulosten analysoinnista. [10]. Tutkimusmenetelmän valintaan vaikuttavat tutkimuskonteksti ja -kysymykset. Esimerkiksi jos referenssivideo on saatavilla tai jos halutaan tutkia vääristymien suuruutta eikä niiden olemassaoloa, ei testimenetelmä välttämättä ole sama [44].

ITU-R BT-500 [21] ja ITU-T P.910 [25] ovat standardeja, joissa määritellään menetelmiä subjektiiviseen videon laadun arviointiin. Niissä esitellään muun muassa DSCQS (Double Stimulus Continuous Quality Scale) [21], DSIS (Double Stimulus Impairment Scale) [21], ACR (Absolute Category Rating) [25], PC (Pair Comparison) [25] ja SSCQE (Single Stimulus Continuous Quality Evaluation) [21] -menetelmät.

DSCQS-menetelmässä vääristynyt video ja referenssivideo esitetään arvioijalle satunnaisessa järjestyksessä. Molempien videoiden arviointi tapahtuu jatkuvalla asteikolla. Havainnoitsijalle ei kerrota, missä järjestyksessä videot esitetään. Arvioijat voivat testistä riippuen vaihdella vapaasti näytettävää videota, tai videot voidaan näyttää järjestelmällisesti peräjälkeen. Jälkimmäisessä tapauksessa videot näytetään tyypillisesti kahteen kertaan, jolloin toisen näyttökerran aikana tai sen jälkeen tapahtuu molempien videoiden arviointi. DSCQS-menetelmä soveltuu esimerkiksi jonkin systeemin aiheuttamien vääristymien suuruuden tutkimiseen, kun referenssivideo on saatavilla.

DSIS-menetelmä on samankaltainen kuin DSCQS-menetelmä, mutta siinä arvioijat ovat tietoisia, missä järjestyksessä vääristynyt video ja referenssivideo näytetään. Havainnoitsijat arvioivat vääristyneen videon suhteessa referenssivideoon vääristyneen videon esityksen aikana tai sen jälkeen diskreetillä viisiasteisella arviointiasteikolla. ITU-T P.910:n vastaava menetelmä on DCR (Degradation Category Rating).

ACR-menetelmässä sekvenssit esitetään yksitellen ja jokaisen esityksen jälkeen arvioidaan kyseinen sekvenssi. Arviointi suoritetaan perusversiossa diskreetillä viisiasteisella arviointi-asteikolla, mutta myös esimerkiksi enemmän erotteleva yhdeksänportainen asteikko on mahdollinen. ACR on nopea menetelmä ja sitä kannattaa käyttää, jos näytteitä on paljon [33]. ITU-R BT-500:n vastaava menetelmä on SS (Single Stimulus).

PC:ssä eli parivertailussa saman videokontentin eri versioita verrataan keskenään. Tässä havainnoitsija päättää kahden version esittämisen jälkeen, kumpi niistä oli parempi. Menetelmässä testataan kaikki mahdolliset versio-kombinaatiot ainakin kerran. PC:tä pidetään luotettavana ja yksinkertaisena menetelmänä. Se ei ole kuitenkaan soveltuvainen isolle määrälle näytteitä, koska vertailupareja tulisi tällöin liikaa.

SSCQE-menetelmässä videon laatua arvioidaan reaaliaikaisesti sekvenssin aikana jatkuva-asteikkoisella liukusäätimellä. SSCQE on hyödyllinen, jos videon laatu vaihtelee paljon sekvenssin aikana, ja halutaan selvittää videon laatua ajan funktiona.

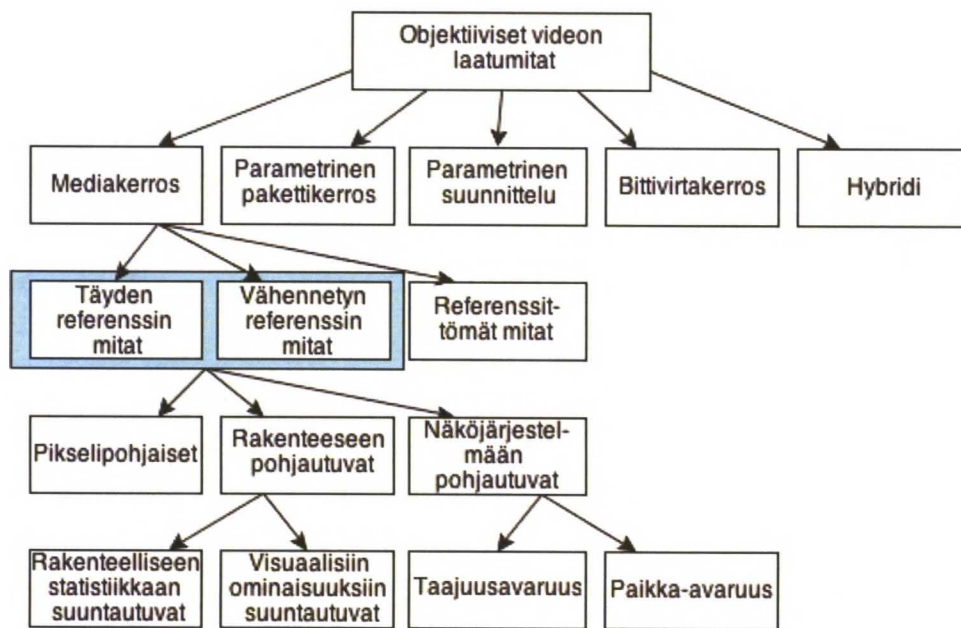
3.2 Objektiiviset mitat

3.2.1 Objektiivisten mittojen luokittelu

Yleinen tapa on jakaa objektiiviset videon laatumitat täyden referenssin, vähennetyn referenssin ja referenssittömiin mittoihin [23]. Täyden referenssin mitat käyttävät laskennassa hyödyksi alkuperäistä vääristymätöntä referenssivideota, johon systeemin läpi mennyttä videota verrataan. Vähennetyn

referenssin mitoissa vain osaa referenssivideon informaatiosta käytetään. Referenssittömät mitat puolestaan eivät käytä mitään informaatiota referenssivideosta. Takahashi ym. [57] kutsuvat näitä menetelmiä mediakerrosmitoiksi. Niissä käytetään syötteenä ainoastaan videokuvia, eikä esimerkiksi videopakettien otsakeinformaatioon, siirtoverkkoon tai päätteisiin liittyviä parametreja.

Chikkerur ym. [8] jakavat täyden referenssin ja vähennetyn referenssin mitat edelleen perinteisiin pikselipohjaisiin (engl. traditional point-based), kuvan rakenteeseen (engl. natural visual characteristics) pohjautuviin ja ihmisen näköjärjestelmään (engl. human visual system) pohjautuviin mittoihin. Kuvan rakenteeseen pohjautuvat mitat jakautuvat edelleen rakenteelliseen statistiikkaan (engl. natural visual statistics) ja visuaalisiin ominaisuuksiin (engl. natural visual features) suuntautuviin mittoihin. Ihmisen näköjärjestelmään pohjautuvat mitat jakautuvat paikka- ja taajuusavaruusmittoihin. Objektiivisten videon laatumittojen luokittelu on nähtävissä kokonaisuudessaan kuvassa 3.1.



Kuva 3.1: Objektiivisten videon laatumittojen luokittelu [8].

Pikselipohjaiset mitat eivät ota huomioon videon rakennetta tai ihmisen havainnointikykyä, vaan ne muodostetaan laskutoimitusten avulla suoraan kuvien pikseleiden arvoista. Näitä mittoja ovat muun muassa MSE (Mean Squared Error) ja PSNR (Peak-Signal-to-Noise Ratio) [8]. MSE on

neliöllinen keskiarvovirhe, joka lasketaan ottamalla keskiarvo referenssi- ja vääristyneen kuvan pikselikohtaisista arvoista [61]. PSNR on puolestaan MSE:stä laskettava mitta, joka ottaa huomioon myös pikselien intensiteettijakauman [61]. PSNR mahdollistaa eri dynaamisen alueen sisältävien kuvien vertailun.

Videon **rakenteelliseen statistiikkaan** suuntautuvat mitat hyödyntävät tilastollisia menetelmiä kuten keskiarvoa, varianssia ja jakaumia videon laadun mittauksessa [8]. Tällainen videon laatumitta on esimerkiksi VSSIM (Video Structural SIMilarity) [62]. VSSIM perustuu tunnettuun SSIM-kuvanlaatumittaan (Structural SIMilarity), joka laskee keskiarvojen, keskijajontojen ja kovarianssien avulla kuvan laatuarvion. VSSIM-mitta laskee videon arvon lokaali-, kuva- ja sekvenssitasolla. Lokaalilaatu saadaan SSIM:n funktion avulla soveltamalla sitä kolmeen värikanavaan. Värikanavista saatuja arvoja painotetaan eri määrä yksittäisen kuvan arvon muodostuksessa. Kuva-arvoja painotetaan edelleen sekvenssitasolla, jolloin lopputuloksena on arvio koko videosekvenssille.

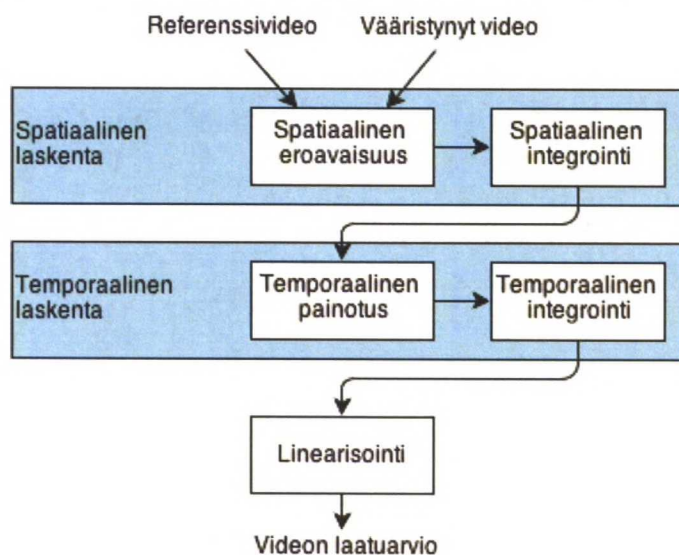
Videon **visuaalisiin ominaisuuksiin** suuntautuvat mitat käyttävät hyödykseen muun muassa mittoja videon sumentuneisuudesta sekä laatikoitumisesta ja kuvien segmentointia löytääkseen videosta merkittäviä visuaalisia piirteitä. Lisäksi mitat voivat yrittää etsiä kuvan reunakohtia, mitkä mahdollistavat kuvan eri alueiden tunnistamisen [8]. Eräs tällainen mitta on VQM (Visual Quality Metric) [45], joka mittaa muun muassa videon sumentuneisuutta, liikkeen laatua, laatikoitumista, värejä, terävöityksen määrää ja kohinan tasoa niin spatiaalisesti kuin temporaalisesti.

Ihmisen **näköjärjestelmään** pohjautuvat mitat pyrkivät mallintamaan ihmisen näköjärjestelmän muodostamaa havaintokokemusta videon laatuarvion muodostuksessa. Taajuusavaruusmitoissa vääristymien suuruutta eri taajuusalueissa mitataan esimerkiksi diskreettien kosinimuunnosten ja aalokemuunnosten avulla. Yksi taajuusavaruuteen luokiteltu mitta on MOVIE (MOtion-based Video Integrity Evaluation) [53]. Siinä muun muassa Gabor-suodattimien avulla simuloidaan ihmisen näköaivokuoren ominaisuuksia.

Paikka-avaruudessa mitataan vääristymien suuruutta esimerkiksi lokaalien gradienttien muutosten tai havainnollisesti merkittävien kuvan ominaisuuksien avulla [8]. Yksi paikka-avaruuteen luokiteltava mitta on PVQM (Perceptual Video Quality Metric) [17]. Se muodostaa videon laatuarvion videon reunakohtien, temporaalisen vaihtelevuuden ja värivirheiden määrien pohjalta.

3.2.2 Täyden referenssin videon laatumittojen laske- minen

Barkowsky ym. [5] esittävät, että täyden referenssin videon laatumittojen laske-
minen koostuu tyypillisesti kolmesta vaiheesta: spatiaalisesta ja temporaal-
lisesta laskennasta sekä linearisoinnista. Spatiaalinen laskentaan sisältyvät
vaiheet ovat spatiaalinen eroavaisuus ja spatiaalinen integrointi. Temporaali-
seen laskentaan puolestaan kuuluvat temporaalinen painotus ja temporaali-
nen integrointi. Lisäksi lopuksi usein suoritetaan mittatuloksien linearisointi.
Tämä rakenne on esitetty kuvassa 3.2. Kuvan kaikkia vaiheita ei välttämättä
esiinny kaikissa laatumitoissa tai ne saattavat olla erittäin yksinkertaisia.



Kuva 3.2: Videon laatuarvion muodostuminen täyden referenssin mitoissa [5].

Spatiaalisessa eroavaisuudessa lasketaan eroja alkuperäisen ja vää-
ristyneen kuvan välillä. Tämän tuloksena muodostuu yleensä kuvakohtainen
vääristymäkartta, jossa esitetään pikselikohtaiset vääristymät. Tässä vaihees-
sa voidaan painottaa enemmän kuvan kohtia, joissa informaatiota on kai-
kista eniten [63]. Esimerkiksi kuvan korkeaenergisten kohtien löytäminen
voi olla kannattavaa, koska ne todennäköisesti herättävät eniten visuaa-
listaa kiinnostusta katsojassa. Kuvan kohtien painottaminen voi perustua
myös silmänliikedataan tai malleihin, jotka tunnistavat kuvasta silmiin-
pistävimmät objektit [1].

Spatiaalisessa integroinnissa vääristymäkartasta lasketaan kuvalle yk-
sittäinen arvo tai muutamia arvoja. Useimmiten tässä lasketaan kuvalle kes-

kiarvo kaikista vääristymäkartan pikseleistä.

Videon laatu saattaa vaihdella merkittävästi videonäytteen aikana. Tästä syystä videota on perusteltua tarkastella myös temporaalisesti. **Temporaalisessa painotuksessa** sekvenssin kuvat saavat painokertoimen. **Temporaalisessa integroinnissa** muodostetaan arvo koko videosekvenssille. Tämä onnistuu yksinkertaisimmillaan laskemalla keskiarvo tai euklidinen normi videosekvenssin painotettujen kuvien arvoista.

Linearisointi on viimeinen vaihe. Siinä temporaa-lisen integroinnin arvo on tarkoitus muuntaa johonkin mielekkääseen arvoon. Tämän arvon on tarkoitus muuttua lineaarisesti suhteessa videon havaittuun laatuun. Linearisoinnissa arvot voidaan kuvata vastaamaan subjektiivisia laatu-arvioita. Vaiheen jälkeen lopullinen videon laatu-arvio on saatavilla.

3.3 Temporaa-linen videon laadun arviointi

Yksinkertaisin menetelmä videon laadun temporaaliseen laskentaan on keskiarvoistaminen. Siinä painotetaan videosekvenssin jokaista kuvaa yhtä paljon kaavalla [49]:

$$\bar{X} = \frac{1}{T} \sum_{t=1}^T OM(t), \quad (3.1)$$

jossa T on kuvien määrä videosekvenssissä ja $OM(t)$ on spatiaalisessa laskennassa saatu laatu-arvo kuvalle t . Jos jokaista videon kuvaa painotetaan yhtä paljon, ei laadun arviossa välttämättä saavuteta riittävää tarkkuutta. Eniten vääristyneillä kuvilla on usein suurin merkitys videon laatu-arviossa. Tätä voidaan simuloida Minkowskin summauksella [50]:

$$MS = \left(\frac{1}{T} \sum_{t=1}^T OM^p(t) \right)^{1/p}. \quad (3.2)$$

Kaava pohjautuu keskiarvon laskemiseen, mutta siihen on lisätty kaksi termiä, joissa esiintyy muuttuja p , joka on Minkowskin eksponentti. Sillä voidaan vaikuttaa siihen, kuinka suuren painoarvon vääristyneimmät kuvat saavat.

Edelliset menetelmät eivät ota huomioon kuitenkaan luvussa 2.2.5 esitettyä muistiefektiä, jota on ilmennyt muun muassa Seferidis ym. [51], Handsin ja Avonsin [15], Zinkin ym. [68], Liun ym. [35] ja Aldridge ym. [4] tutkimuksissa.

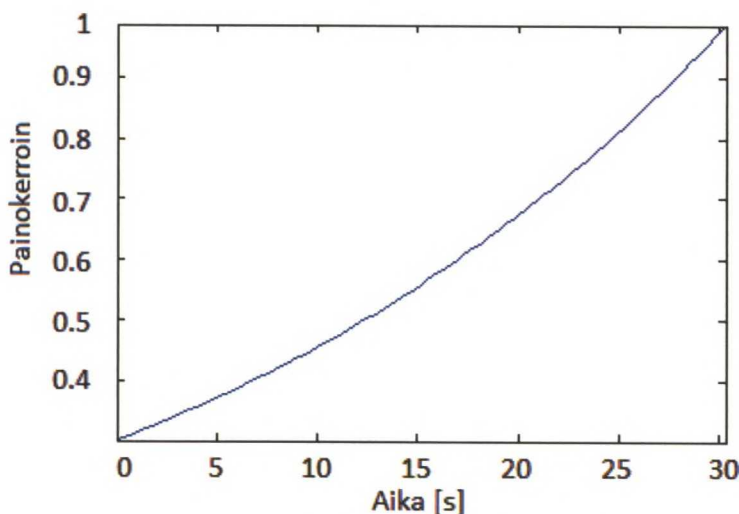
Seferidis ym. ovat todenneet, että 10 sekunnin vääristynyt sekvenssi arvioidaan huonompana kuin 30 sekunnin sekvenssi, joka alkaa samaisella 10

sekunnin vääristyneellä pätkällä ja päättyy 20 sekunnin vääristymättömään pätkään. Heidän mukaansa tämä johtuu siitä, että vääristyneen sekvenssin aiheuttama havainto alkaa lieventyä ja unohtua vääristymättömän sekvenssin aikana.

Hands ja Avons ovat vertailleet tilanteita, joissa 30 sekunnin videosekvenssin alussa tai lopussa oli 5 sekunnin vääristynyt kohta. Tulokset osoittivat, että jos vääristynyt kohta oli videon lopussa, videon laatu oli arvioitu huonompana kuin jos vääristynyt kohta oli videon alussa. Samaa tulokseen ovat päätyneet myös muun muassa Zink ym., joiden tutkimuksessa tosin sekvenssit olivat 10 sekuntia pitkiä ja niissä oli 5 sekunnin vääristymä joko alussa tai lopussa.

Liu ym. ovat vertailleet tilanteita, joissa sekunnin vääristynyt kohta sijaitsi 40 sekunnin sekvenssin alussa, keskellä tai lopussa. Tutkimuksen perusteella jos vääristymä oli pidemmällä kuin 15 sekunnin päässä videosekvenssin lopusta, sen sijainti ei enää vaikuttanut kokonaislaatuarvioihin. Samankaltaiseen tulokseen ovat päätyneet myös Aldridge ym. Heidän mukaan 20–30 sekunnin päässä sekvenssin lopusta sijaitsevat vääristymät painottavat enää vähän tai ei ollenkaan videon kokonaislaatua.

Muistiefektiä voi mallintaa laskemalla keskiarvon vain sekvenssin viimeisistä kuvista [49]. Parempaan tarkkuuteen päästään painottamalla videosekvenssin kuvia eksponenttifunktiolla [5], joka on esitetty kuvassa 3.3, ja tämän jälkeen suorittamalla temporaalisen integroinnin.



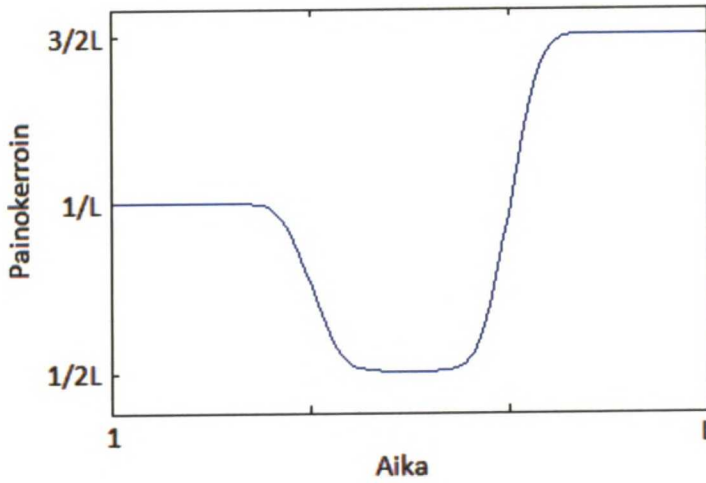
Kuva 3.3: Temporaalinen painotusfunktio 30 sekunnin pituisille videosekvensseille [5].

Muistiefektin mallinnus onnistuu myös eksponentilla painotetulla Minkowskin summauksella [50]:

$$EWMS = \left(\frac{1}{T} \sum_{t=1}^T \exp \left(\frac{t-T}{\tau} \right) OM^p(t) \right)^{1/p}. \quad (3.3)$$

Kaavassa termeillä T , t , p ja $OM(t)$ on sama merkitys kuin kaavassa 3.2. Lisätyssä eksponenttitermissä on muuttuja τ , jolla kontrolloidaan muistiefektin voimakkuutta.

You ym. [67] ovat myös esittäneet videon temporaalista laatua painotavan funktion. Edellisten menetelmien tapaan funktio antaa suuremman painoarvon videosekvenssin lopussa oleville kuville. Aiemmistä menetelmistä poiketen videosekvenssin alkua on kuitenkin painotettu enemmän kuin sekvenssin keskivaihetta. Alku ei kuitenkaan ole saanut niin suurta painoarvoa kuin sekvenssin loppu. Painotusfunktio on esitetty kuvassa 3.4.



Kuva 3.4: Temporaalinen painotusfunktio [67]. L on videon pituus.

Videosekvenssin kuvan k painokerroin on määritetty kaavalla:

$$F(k) = \begin{cases} \frac{1}{L}, & k \leq \frac{L}{3} \\ \frac{1}{2L}, & \frac{L}{3} < k < \frac{2L}{3} \\ \frac{3}{2L}, & k \geq \frac{2L}{3} \end{cases}, \quad (3.4)$$

missä L on kuvien lukumäärä videossa. Funktiota on suodatettu useita kertoja Gaussian-suotimella, jonka seurauksena sen portaittainen rakenne on muuttunut sulavaan muotoon.

Hands [14] on selvittänyt uudemmassa tutkimuksessa, että muistiefektin pituuteen vaikuttaa vääristymän sijainnin lisäksi vääristymän voimakkuus. Hands vertaili sekvenssejä, joiden alussa oli 10 sekuntia joko matalasti tai korkeasti vääristynyt osuus. Tätä seurasi aina joko 0, 30 tai 60 sekunnin vääristymätön osa. Hands totesi, että videon havaittuun laatuun vaikutti pidempään ja voimakkaammin se, jos sekvenssin alku oli erittäin vääristynyt kuin jos se oli vain hiukan vääristynyt.

Luku 4

Videopankit

Tässä luvussa esitetään vapaasti jaettavien videotankkien yleispiirteitä sekä videoita karakterisoivia mitta-arvoja.

4.1 Videotankkien piirteet

Videotankkit ovat videoita sisältäviä kokoelmia. Videon laatuututkimukseen tarkoitettujen videotankkien yhteydessä on usein saatavilla koehenkilöiden subjektiiviset videokohdaiset laatuarviot. Nämä arviot mahdollistavat objektiivisten videon laatua mittaavien algoritmien hyvyyden testaamisen suhteessa havaittuun videon laatuun sekä suhteessa muihin algoritmeihin. [59]

Winklerin tutkimuksessa [66] analysoitiin vapaasti saatavilla olevia videotankkeja. Mukana olivat LIVE (LIVE Video Quality Database), EPFL/PoliMI (EPFL/PoliMI Video Quality Assessment Database), IVC-1080i (IRCCyN/IVC 1080i Database), IVC-ROI (IRCCyN/IVC SD RoI Database), IVP (IVP Database), MMSP-3D (MMSP 3D Video Quality Assessment Database), MMSP-SVD (MMSP Scalable Video Database), NYU-1, NYU-2, NYU-3 (Poly@NYU Video Quality Database), NYU-PL (Poly@NYU Packet Loss Database), VQEG-FR (Video Quality Experts Groups FR-TV Phase I Database) ja VQEG-HD (VQEG HDTV Database). Taulukossa 4.1 on esitetty näiden videotankkien ominaisuuksia. Arvot ovat peräisin artikkelista [66], josta löytyvät myös alkuperäiset viitteet videotankkeihin. Videon pituudet taulukkoon on selvitetty näiden viitteiden pohjalta.

Taulukosta on huomattavissa, että tyypillinen videotankki sisältää noin 10 videokontenttia, joista jokaisesta on luotu noin 10 erilaista versiota ja ne ovat noin 10 sekuntia pitkiä. Poikkeuksena ovat VQEG-HD-videotankki, jossa on paljon suurempi määrä videokontentteja ja versiota; sekä NYU-PL-videotankki, jonka videot ovat vain 2 sekuntia pitkiä. Usein jokaisesta ver-

Taulukko 4.1: Yhteenvedo videopankeista [66].

Videopankki	Vuosi	Kont.	Ver.	Resoluutio	Fps	Pituus [s]
EPFL/PoliMI	2009	12	12	352*288/704*576	25/30	10
IVC-1080i	2008	24	7	1920*1080	25	9–12
IVC-ROI	2009	6	14	720*576	25	8–10
IVP	2011	10	10–14	1920*1080	25	10
LIVE	2010	10	15	768*432	25/50	8,68/10
MMSP-3D	2010	6	5	1920*1080	25	10
MMSP-SVD	2010	3	vaiht.	1280*720	50	10
NYU-1	2008	6	5	352*288/176*144	30	10
NYU-2	2009	4	16	352*288/176*144	30	10
NYU-3	2010	6	15	352*288/176*144	30	10
NYU-PL	2007	17	1	320*240	10–15	2
VQEG-FR	2000	20	16	704*480/720*576	25/30	8
VQEG-HD	2010	49	75	1920*1080	35/30	10
Kont. = Videokontenttien määrä						
Ver. = Versioiden määrä						
Fps = Kuvia sekunnissa						

siosta on saatavilla myös subjektiiviset arviot. Käytetty resoluutio vaihtelee videopankeittain. Enimmillään se on 1920*1080 ja vähimmillään 320*240 pikseliä. Useimpien videopankkien videoiden kuvanopeus on joko 25 tai 30 kuvaa sekunnissa ja videot ovat saatavilla pakkaamattomassa muodossa.

Videoiden versioihin on muodostettu erilaisia vääristymiä riippuen videopankista. Esimerkiksi LIVE-videopankin [54, 55] videokontentteihin on generoitu H.264- ja MPEG-2-pakkauksessa sekä IP- ja langattomissa verkoissa vaikuttavia vääristymiä. Puolestaan VQEG-FR-videopankin [58] videoiden versioissa vaihtelee videoiden bittivirran nopeus, resoluutio ja käytetty pakkausmenetelmä.

Videopankeissa on luonnollisesti käytetty myös erityylisiä ja eriperustein kuvattuja videoita. Esimerkiksi LIVE-videopankki [54, 55] sisältää luonnollisia, ei animoituja ja ei tekstuuripohjaisia videoita, joissa on joko globaalia tai lokaalia liikettä. VQEG-FR-videopankin [58] videokontentteihin on pyritty sisällyttämään vaihtelevasti temporaalista ja spatiaalista informaatiota sekä värejä. Molempien videopankkien useissa videoissa kuvauskohteen rajaus muuttuu kesken videon tai kuvauskohde muuttuu kokonaan. VQEG-videopankin videoissa esiintyy jopa leikkauskohtia.

4.2 Mitta-arvot videoiden karakterisointiin

Videon subjektiiviseen laadun arviointiin keskittyvissä standardeissa [20, 21] esitetään, että spatiaalinen (SI) ja temporaalinen (TI) informaatio ovat oleellisia parametreja videon karakterisoinnissa. Näiden parametrien avulla voidaan ennustaa vääristymien havaittavuus. Videokontentit suositellaan valittavaksi videon laatututkimukseen niin, että ne kattavat mahdollisimman suuren alueen spatiaalisessa ja temporaalisessa avaruudessa.

Standardissa [20] esitetty SI-mitta indikoi videon spatiaalisten yksityiskohtien määrää, ja se lasketaan kaavalla:

$$SI = \max_{time} \{std_{space}[Sobel(F_n)]\}. \quad (4.1)$$

Kaavassa videon luminanssikanavan jokainen kuva suodatetaan Sobel-suotimella, $[Sobel(F_n)]$. Tämän jälkeen jokaisen Sobel-suodatetun kuvan keskihajonta, std_{space} , lasketaan. Videon SI-arvo saadaan valitsemalla suurin arvo, \max_{time} , lasketuista keskihajonnoista.

Standardissa [20] esitetty TI-mitta indikoi videossa olevaa liikettä, ja se lasketaan kaavalla:

$$TI = \max_{time} \{std_{space}[M_n(i, j)]\}. \quad (4.2)$$

Kaava on SI-mittaan verrattuna samankaltainen. Ainoastaan termi $[M_n(i, j)]$ on poikkeava. Se on sekvenssissä kahden peräkkäisen kuvan pikselikohtainen eroavaisuus. Siinä i viittaa kuvan pikseliriveihin ja j sarakkeisiin.

Winklerin videopankkitutkimuksessa on vertailtu videopankkien alkuperäisten videoiden mitta-arvoja sekä videoiden vääristyneiden versioiden ja subjektiivisten mitta-arvojen tilastollisia ominaisuuksia. Winkler [66] on esittänyt vaihtoehtoiset laskentakaavat SI- ja TI-mitoille. Kaavat ovat samoihin menetelmiin pohjautuvia kuin edelliset kaavat, mutta ovat astetta monimutkaisempia, koska ne ottavat huomioon muun muassa mittojen resoluutiiriippuvuuden ja videoiden kuvanopeuden. Lisäksi ne laskevat keskiarvon kaikkien videokuvien yli, eivätkä käytä vain maksimiarvoa. Winkler esittää myös videoita karakterisoivan väriparametrin (CI).

Winklerin SI-mitta [66] saa suuria arvoja, jos videossa on paljon yksityiskohtia. Se pohjautuu niin ikään Sobel-suotimiin, joilla suodatetaan videon kuvia. Niillä pystytään tunnistamaan kuvissa reunakohtia. Tässä Sobel-suodatus (s_r) koostuu kahdesta suodatuksesta (s_v , s_h), jotka pystyvät tunnistamaan joko vaaka- tai pystysuuntaisia reunoja kaavalla: $s_r = \sqrt{s_v^2 + s_h^2}$. s_v ja s_h saadaan laskettua kaavoilla:

$$s_v = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * A \quad ja \quad s_h = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * A. \quad (4.3)$$

Kaavoissa $*$ on konvoluutio-operaattori ja A -matriisi on kuvan luminanssi-kanava. Lopullinen kuvan SI-mitta saadaan kaavalla:

$$SI = \sqrt{\frac{L}{1080}} \sqrt{\sum_{r=1}^P \frac{s_r^2}{P}}. \quad (4.4)$$

Kaavassa P on pikselien määrä kuvassa ja L on kuvan vertikaalinen resoluutio. Termillä $\sqrt{\frac{L}{1080}}$ pyritään vähentämään SI-mitan resoluutioriippuvuutta. Koko videon SI-arvo saadaan laskemalla keskiarvo yksittäisten kuvien SI-arvoista.

Winklerin TI-mitta [66] lasketaan kaavalla:

$$TI = \frac{f}{L} \sqrt{\sum_{r=1}^M \frac{|v_r|^2}{M}}. \quad (4.5)$$

Kaavassa M on blokkien määrä videossa. Kahden kuvan blokkien välillä on liikevektori, v . Kuvien vastinblokit on sovitettu SAD-menetelmällä¹ (Sum of Absolute Differences). Merkinnällä $|v|$ tarkoitetaan vektorin pituutta. L on videon vertikaalinen resoluutio ja f on videon yksittäisen kuvan esitys aika. Normalisointitermin $\frac{f}{L}$ avulla pyritään siihen, että mitta on vertailukelpoinen eri resoluution ja kuvataajuuden sisältävien videoiden välillä.

CI-mitta, jota Winkler käytti videopankkianalysissään, lasketaan kaavalla [16]:

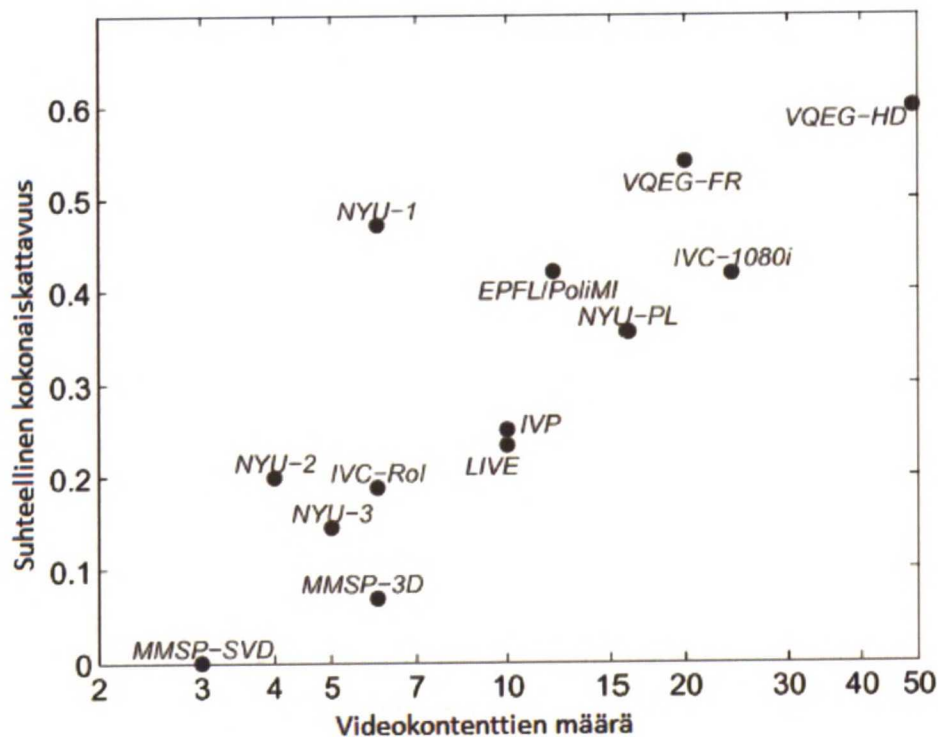
$$CI = \sqrt{\sigma_{RG}^2 + \sigma_{YB}^2} + 0.3 \sqrt{\mu_{RG}^2 + \mu_{YB}^2}. \quad (4.6)$$

Kaavassa σ on keskihajonta ja μ on keskiarvo. Kaavoilla $RG=R-G$ ja $YB=0,5(R-G)-B$ saadaan laskettua CI-mittaan tarvittavat väriavaruudet. Esimerkiksi μ_{RG}^2 tarkoittaa RG-väriavaruudessa kuvan pikselien keskiarvoa toiseen potenssiin korotettuna. Koko videon CI-arvo lasketaan ottamalla keskiarvo yksittäisten kuvien CI-arvoista. Värimitta indikoi värien vaihtelevuutta ja intensiteettiä kuvassa.

Winklerin [66] esittämä yksittäinen suure, jolla videopankkeja voi verrata, on suhteellinen kokonaiskattavuus (relative total coverage). Winkler on laskenut suhteellisen kokonaiskattavuuden taulukossa 4.1 esitetyille videopankeille. Se saadaan ottamalla kuutiojuuri normalisoidusta SI-CI-TI-avaruuden

¹<http://avisynth.org.ru/mvtools/mvtools2.html>

tilavuudesta, jonka videoiden mitta-arvot kattavat. Kuvassa 4.1 on kopio Winklerin artikkelissa [66] olevasta kuvasta, jossa videopankkien suhteellinen kokonaiskattavuus on esitetty videokontenttien määrän funktiona. Siitä on nähtävillä, että tyypillisesti enemmän videoita sisältävät videopankit ovat saaneet suurempia suhteellisen kokonaiskattavuuden arvoja.



Kuva 4.1: Videopankkien suhteellinen kokonaiskattavuus [66].

Luku 5

Tutkimusmenetelmät

Tutkimuksen tavoitteena oli määrittää, miten videon pituus vaikuttaa kokonaislaatu-arvioon, kun vääristymä oli sijoitettu joko sekvenssin alkuun, keskelle tai loppuun. Aiemmat tutkimukset osoittavat, että videon loppuosan laatu vaikuttaa eniten laatu-arvioon. Aiemmissa tutkimuksissa ei ole kuitenkaan tutkittu sitä, miten videon pituus ja vääristymän sijainti vuorovaikuttavat.

Tässä luvussa kerrotaan tutkimusta varten tuotetuista videomateriaaleista ja suoritetuista koehenkilötesteistä. Luvussa 5.1 kerrotaan videomateriaaleista, joita valmistettiin testejä varten ja luvussa 5.2 kerrotaan testimenetelmästä, jota käytettiin koehenkilötesteissä. Luku 5.3 sisältää kuvaukset kahdesta esitestistä, joiden perusteella valittiin videokontentit sekä määriteltiin videoille vääristymätasot varsinaiseen testiin. Varsinaiset testit I ja II on esitetty luvuissa 5.4 ja 5.5.

Testissä I vääristymän pituus oli vakio, 5 sekuntia. Testi II oli muuten vastaava kuin testi I, mutta siinä videoissa olleiden vääristymien pituus vaihteli suhteellisesti ollen kaikissa tilanteissa aina puolet koko sekvenssin pituudesta. Testi I on nimetty ”vakiopituinen vääristymä” ja testi II on nimetty ”suhteellinen vääristymä”.

5.1 Testivideot

Luvussa 4.1 esiteltiin vapaasti jaettavien videopankkien ominaisuuksia. Näitä videopankkeja ei voitu soveltaa tässä tutkimuksessa, koska tarve oli videosekvensseille, jotka ovat tarpeeksi pitkiä ja joiden sisältö ei muutu merkittävästi sekvenssin aikana. Useimpien videopankkien videoilla on vakiopituus, joka on vain noin 10 s. Näin lyhytkestoiset videot eivät mahdollista sekvenssin pituuden vaikutuksen tutkimista videonlaatu-arviossa. Sisällön vaihtelevuus

sekvenssin aikana ei myös ole toivottua, koska luvussa 2.2.5 todettiin videosisällön vaikuttavan arviointeihin. Luvussa 4.2 esitettyjen mitta-arvojen muutos kuvien välillä pidettiin minimissä ja kuvauskohdetta tai kuvakulmaa ei vaihdettu sekvenssin aikana, millä pyrittiin minimoimaan videokontentin vaikutus arvioinneissa. Muuten eripituiset sekvenssit samasta videokontentista eivät olisi vertailukelpoisia.

Tutkimusta varten luodun videopankin videot kuvattiin Canon EOS 5D mark II -kameralla¹, joka pakkaa videon häviöllisesti H.264-tekniikalla käyttäen Baseline-profilia. Kuvakaappaukset videoista on esitettynä kuvassa 5.1. Alkuperäiset videot ovat pituudeltaan 40 sekuntia. Videot on kuvattu täydellä teräväpiirtotarkkuudella (1920*1080) ja niissä on 25 kuvaa sekunnissa. Videoissa ei ole ääntä, koska multimodaalinen äänen ja videon yhteisvaikutuksen tutkiminen rajattiin tutkimuksen ulkopuolelle.

Subjekttiivisen videonlaadun arviointimenetelmiä sisältävien standardien suosituksia [21, 25] noudattaen kuvatessa pyrittiin siihen, että videoiden mitta-arvot kattavat mahdollisimman suuren alueen spatiaali- (SI) ja temporaaliavaruudessa (TI). Lisäksi eri värejä valittiin vaihtelevasti videoihin, koska myös värimittaa (CI) on pidetty videopankkien laadun kriteerinä [66]. Tässä työssä SI-, TI-, ja CI-arvojen laskemiseen on käytetty luvussa 4.2 esitettyjä Winklerin kaavoja, joita hän on käyttänyt omassa tutkimuksessaan. Näin tuotetusta videopankista lasketut mitta-arvot ovat vertailukelpoisia tutkimuksessa [66] laskettuihin mitta-arvoihin.

TI-arvojen laskentaa varten tarvittavat liikevektorit on laskettu samalla tavalla kuin tutkimuksessa [66]: AVISynth² ja VirtualDub³-ohjelmilla. MVTools⁴ on AVISynth:n lisäosa, jonka MAnalyse-funktiolla on mahdollista löytää liikevektoreita videoista ja tallentaa ne raakadatana tiedostoon. Liikevektoreiden parsiminen raakadatasta toteutettiin Matlab:lla luodulla ohjelmalla.

SI- ja TI-arvojen laskennassa on käytetty videokuvien luminanssikanavaa, L. Se on laskettu kaavalla $L=0,299R+0,587G+0,114B$ [66], missä R, G ja B ovat kuvan värikanavakohtaiset matriisit.

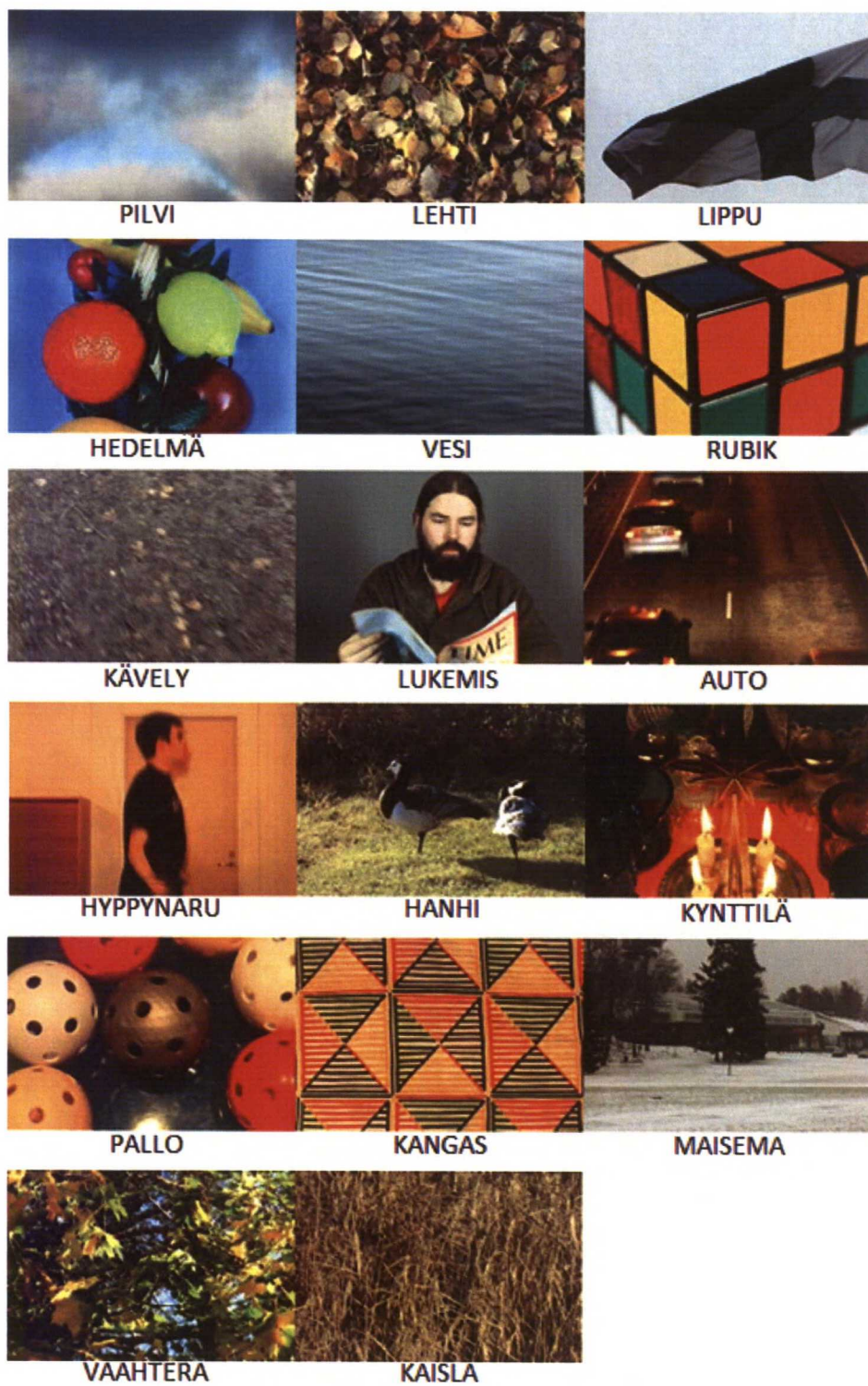
Kuvassa 5.2 on esitetty, miten kuvattujen videoiden mitta-arvot ovat jakautuneet SI-, CI-, ja TI-avaruuksiin. Esimerkiksi yksityiskohtaiset tuulessa liikkuvat kaislat ovat saaneet melko suuret SI- ja TI-arvot verrattuna muihin videoihin, kun puolestaan videon CI-arvot ovat suhteellisen matalat johtuen sen ruskeahkosta värimaailmasta. Toisena esimerkkinä mainittakoon rubik-video, jossa on paljon värejä (CI), mutta ei juurikaan yksityiskohtia (SI), vaan

¹http://www.canon.fi/For_Home/Product_Finder/Cameras/Digital_SLR/EOS_5D_Mark_II

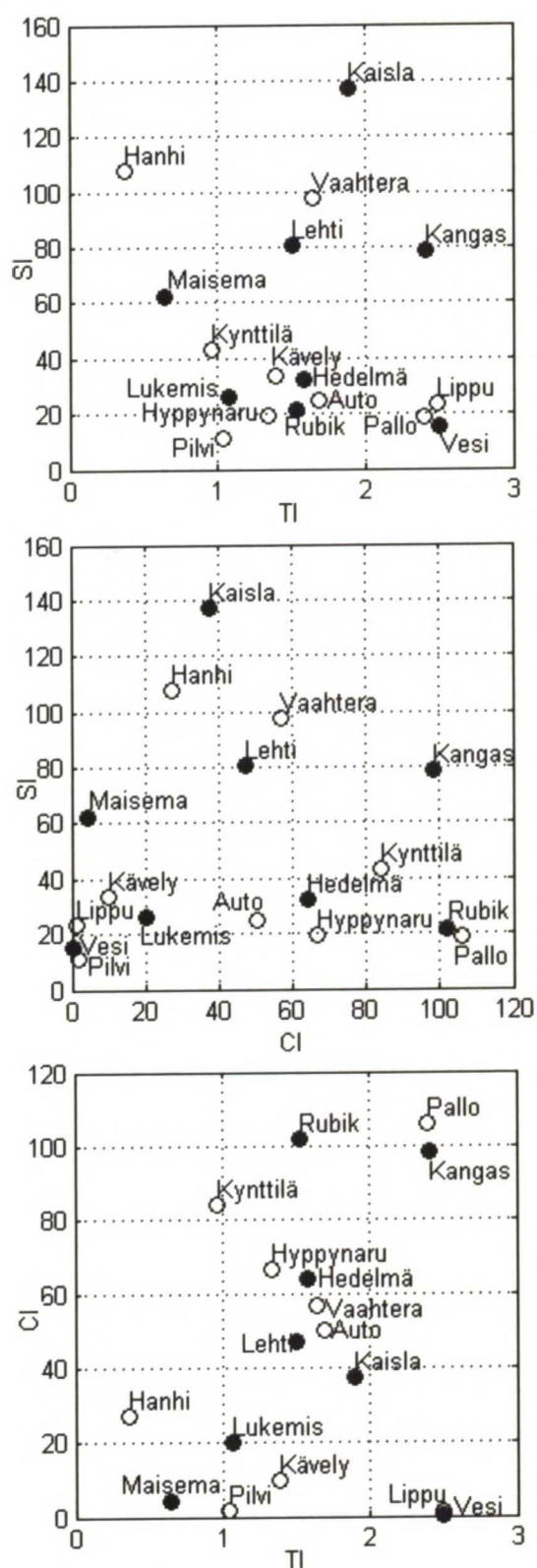
²http://avisynth.nl/index.php/Main_Page

³<http://www.virtualdub.org/>

⁴<http://avisynth.org.ru/mvtools/mvtools.html>



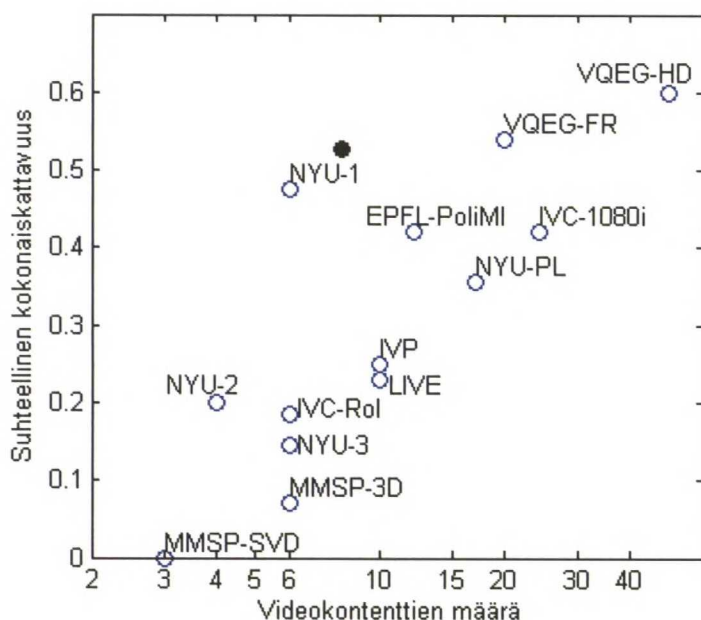
Kuva 5.1: Kuvakaappaukset testivideoista.



Kuva 5.2: Testivideoiden sijainnit SI-TI-, CI-TI-, ja SI-CI-avaruuksissa. Tum-malla värillä on esitetty varsinaisiin testeihin valitut videot.

isoja värillisiä alueita. Kuution hitaan pyörimisliikkeen johdosta liikettä (TI) videossa on keskvierotomäärä verrattuna muihin videoihin.

Yksittäinen suure, jolla videopankkeja voi verrata, on suhteellinen kokonaiskattavuus. Kyseisen suureen arvoksi saatiin 0,62; kun se laskettiin kaikista alkuperäisistä videoista luvussa 4.2 esitetyllä menetelmällä. Kun laskenta suoritettiin varsinaisiin testeihin valituille kahdeksalle videokontenteille saatiin vastaavaksi arvoksi 0,52. Tämä arvo on esitetty tummalla pisteellä kuvassa 5.3, jossa on myös tutkimuksessa [66] mitattujen muiden videopankkien arvot. Tässä tutkimuksessa luodun videopankin suhteellisen kokonaiskattavuuden arvo on kolmanneksi suurin vertailtujen videopankkien joukossa. Tämän perusteella luodun videopankin videot sisältävät yksityiskohtia, liikettä ja värejä lähes yhtä vaihtelevasti kuin kattavimmat muut vapaasti jaettavat videopankit.



Kuva 5.3: Videopankkien suhteellinen kokonaiskattavuus. Luotu videopankki on esitetty tummalla pisteellä. Muiden videopankkien arvopisteet on arvioitu artikkelin [66] kuvan 4 perusteella.

Kuvatessa on pyritty varioimaan sitä, onko videossa globaalia vai lokaalia liikettä sekä kuvausympäristöjä, kuten sisällä ja ulkona kuvaamista. Lisäksi on vaihdeltu sitä, sisältävätkö videot ihmisiä tai muita eläviä olentoja. Suurin osa videoista on kuvattu jalustalta kameraa liikuttamatta, jolloin ainoas-

taan kuvattavan objektin liike on luonut videon tuntua näytteeseen. Näitä videoita ovat pilvi-, lippu-, vesi-, rubik-, lukemis-, auto-, hyppynaru-, hanhi-, kynttilä-, pallot-, kangas-, maisema-, vaahtera-, ja kaislavidet. Näistä tosin objektin liikuttaminen rubik-, pallo-, ja kangasvideoissa saattaa luoda illuusion, että itse kameraa on liikutettu kuvattaessa. Rubik-videossa kuu-tio on pyörivällä alustalla, pallovideossa salibandypalloja heilutetaan vadis-sa ja kangasvideossa kangasta liu'utetaan vaakasuuntaisesti. Käsivaralta ku-vattuja ovat lehti- ja kävelyvideo. Lehtivideossa kuvaaja liikuttaa kameraa ympyränmuotoisesti, ja kävelyvideossa kuvataan polkua kuvaajan kulkiessa eteenpäin. Hedelmävideossa kameraa liikutetaan jalustalla hitaasti rullaa-malla alaspäin.

Yhteensä videoita oli 17. Varsinaisiin testeihin valittiin esitestien perus-teella vain osa videokontenteista, koska muuten testi olisi kestänyt liian kau-an.

5.2 Testimenetelmä

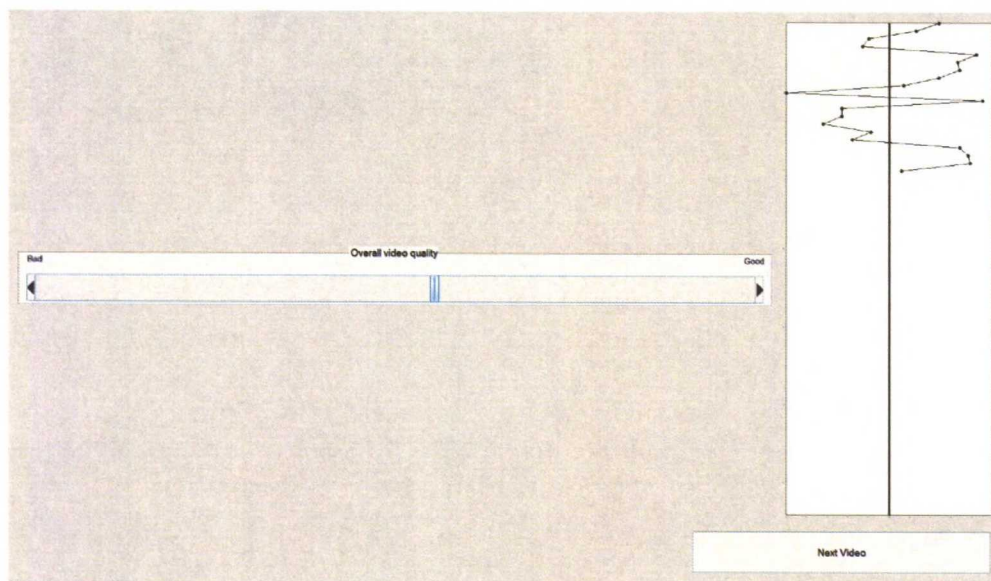
Luvussa 3.1 esiteltiin subjektiiviseen videon laadun arviointiin soveltuvia tut-kimusmenetelmiä. Tähän työhän valittiin käytettäväksi jatkuva-asteikkoinen ACR-arviointimenetelmä [21], koska se mittaa yksinkertaisesti ja nopeasti vi-deon kokonaislaadun sen esityksen jälkeen. Koska videoita oli paljon, ei sel-laista arviointimenetelmää voinut valita, jossa näytetään jokaisen arvioita-van videon yhteydessä vääristymätön referenssivideo. Monien menetelmien, kuten parivertailun, käyttö olisi johtanut liian pitkään testin kestoon. Tar-koitus oli tutkia koettua havaintoa sekvenssin jälkeen eikä jatkuva-aikaista laatuhavaintoa, mikä sulki pois reaaliaikaisen arvioinnin sisältävät testime-netelmät. Jos tutkitaan vääristymän havaittavuutta, standardeissa suositel-laan käytettävän sekvenssejä, jotka ovat pituudeltaan 10-15 s. Tämän työn varsinaisissa testeissä käytettiin kuitenkin standardien suosituksia pidempiä videoita, koska muuten videon pituuden vaikutusta havaittuun laatuun ei ollut mahdollista tutkia. Jatkuva arviontiasteikko mahdollisti yksiselitteisen datan keräämisen, jolloin mahdolliset pienetkin erot videon laadussa saatiin talletettua. ACR-menetelmä mahdollisti myös piilotettujen referenssivideoi-den käytön.

Subjektiivisessa testauksessa käytettiin Helsingin yliopiston Visual Cog-nition -tutkimusryhmän kehittämää VQone-ohjelmaa. Tämä ohjelma on to-teutettu Matlab:lla.

Testin aluksi näytettiin yhdestä videosisällöstä kaksi esimerkkiversio-ta, jotka edustivat huonoa ja hyvää laatua. Kyseiseksi esimerkiksi valit-tiin lippuvideo. Esimerkkivideoiden esityksen jälkeen koehenkilöillä oli tieto

vääristymätyypistä ja ymmärrys siitä, mikä on videoiden laatuvahteluväli. Esimerkkivideo ei sisältynyt enää arvioitavien videoiden joukkoon testissä, paitsi esitestissä I.

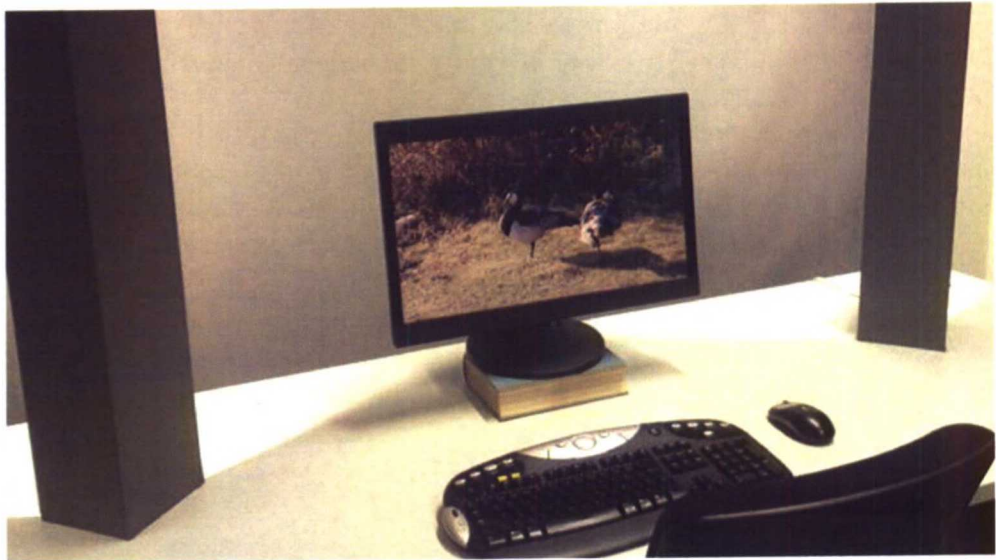
Esimerkkivideoiden jälkeen jokainen testivideo näytettiin satunnaisessa järjestyksessä yksi kerrallaan. Videoiden kokonaislaadun arviointi tapahtui videon katselun jälkeen. Arviointi suoritettiin hyvästä huonoon ulottuvalle arviointiasteikolla (Bad–Good). Koehenkilöt pystyivät näkemään edelliset vastauksensa arviointikäyttöliittymän oikeassa reunassa. Arvioitavia videoita ei ollut mahdollista nähdä uudestaan. Testistä tallentui tiedostoon jokaista videota kohden käyttäjän antaman laatuarvion numeerinen vastine väliltä 0–100. Arviointikäyttöliittymä on nähtävissä kuvassa 5.4.



Kuva 5.4: Arviointikäyttöliittymä.

Kokeet pidettiin Aalto-yliopiston Mediatekniikan laitoksella. Koeasetelma noudatti subjektiivisiin videonarviointitesteihin kohdistuvia suosituksia laboratorioympäristöstä [21]. Kuvassa 5.5 esitetty koeasetelma oli valaistu kahdella pystysuuntaisella loisteputkella (66lux, 5810K). Katselukulma oli suoraan näytön edestä, ja suositeltua [21] noin kahden metrin katseletäisyyttä käytettiin. Koeasetelman 22-tuumainen LCD-näyttö (Eizo ColorEdge) kalibroitiin X-Riten⁵ i1Match-ohjelmalla ja i1Monitor-spektrometrillä tavoitearvoihin: 6500K, 120cd/m² ja gamma-arvoon 2,2.

⁵<http://www.xrite.com>



Kuva 5.5: Koeasetelma.

Testin aluksi koehenkilöiltä kysyttiin taustatietoja (ikä, sukupuoli, ammatti) ja oliko heillä mitään näköön liittyviä oireita, kuten värisokeutta. Taustatietokyselyn jälkeen koehenkilöille suoritettiin vielä kaksi näkötestiä: lähinäkötesti [31] ja värinäkötesti [37].

Ennen esitestejä koehenkilöt lukivat seuraavan ohjeistuksen läpi:

In this experiment, you will see short distorted/blurred video sequences on the screen that is in front of you. At the end of each presentation, you should judge its overall quality by using continuous scale ranging from bad to good. Try to use whole scale when assessing the videos. There are less than 100 videos in total and the test will last about half an hour. Before the test, you will see two example videos which represent bad and good quality. There are no wrong or right answers in this test, we only want to hear your opinion. You don't have to be all the time in the test, but it is recommended to finish the test.

Testin ohjeistus oli englanniksi, mikä mahdollisti monipuolisen koehenkilöotannan. Vaatimus suomenkielen osaamisesta olisi rajannut pois useita koehenkilöitä. Testin ohjeistustekstiin tehtiin muutama muutos varsinaisien testien pidempien videoiden takia, ja siinä se oli kokonaisuudessaan:

In this experiment, you will see short blurred video sequences on the screen that is in front of you. At the end of each presentation, you should judge its overall quality by using continuous scale ranging from bad to good. Try to use whole scale when assessing the videos. There are less than 100 videos in

total and the test will last about 45 minutes. The length of the videos varies between 10 and 40 seconds. Before the test, you will see two example videos representing bad and good quality. There are no wrong or right answers in this test, we only want to hear your opinion.

Jos koehenkilö ei ymmärtänyt ohjeistusta tai siinä oli jotain epäselvää, hänen pyydettiin olemaan yhteydessä kokeenpitäjään. Testin kulku ja arviointikäyttöliittymä käytiin vielä kuitenkin suullisesti läpi jokaisen koehenkilön kanssa.

Testejä mainostettiin uutisryhmien, sähköpostilistojen ja kaverikontaktien kautta. Koehenkilöille jaettiin palkkioksi testiin osallistumisesta elokuvalippuja.

5.3 Materiaalin valinta

Tässä luvussa esitetään esitestit I ja II, joiden perusteella valittiin videokontentit sekä vääristymätasot varsinaiseen testiin.

Esitestien I ja II tavoitteena oli löytää kontenttikohtaiset laatutasot, jotka havaitaan yhtä suurina. Jokaisesta videokontentista oli tarkoitus löytää hyvän ja huonon laadun tasot, joiden välillä varsinaiseen testiin valmistettujen videoiden laatu vaihteli. Niin sanottu hyvän laadun taso oli videonlaatuarvioissa taso, joka saavutettiin, kun videota huononnettiin vain vähän. Jotta sama taso saavutettiin jokaisessa videossa, jouduttiin joitakin videoita huonontamaan hiukan enemmän kuin toisia. Vastaavasti toimittiin myös niin sanotun huonon laadun kohdalla, mutta nyt videoita huononnettiin merkittävästi enemmän kuin hyvän laadun tilanteessa.

Hyvän ja huonon laadun taso pyrittiin asettamaan niin, että niillä oli selkeä ero. Huono laatu tuli asettaa tasolle, joka eroaa tilastollisesti merkittävästi hyvän laadun tasosta. Hyväksi laaduksi haluttiin täysin vääristymättömän alkuperäisen version sijaan hiukan huononnettu versio, koska eri videokontenttien täysin vääristymättömät versiot eivät ole laadultaan samantasoisia. Esimerkiksi heikot valaistusolosuhteet tai heiluva kamera huonontavat videon laatua jo lähtökohtaisesti.

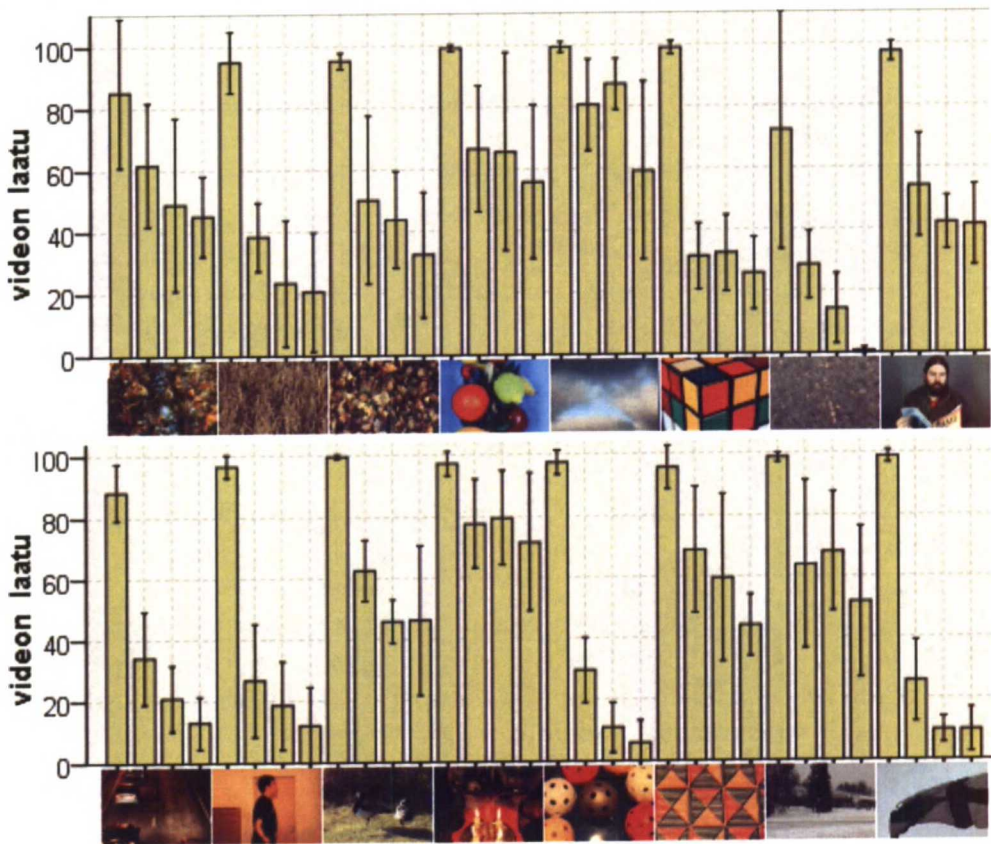
5.3.1 Esitesti I

Esitesti I oli alun perin Visuaalisen median laadun⁶ -kurssille tehty harjoitustyö. Esitestin I tavoite oli löytää videoihin yhtenäisiä laatutasoja pudottamalla videosekvensseistä kuvia simuloiden näin verkoissa tapahtuvia siir-

⁶<https://noppa.aalto.fi/noppa/kurssi/t-75.5200/>

tohäiriöitä. Kaikkien videoiden pituudeksi leikattiin 10 sekuntia, ja jokaisesta videosta luotiin neljä erilaatuista versiota. Testimenetelmä on kuvattu luvussa 5.2. Esitestissä I käytettiin kaikkia tuotettuja videokontentteja, joista esimerkkinä toimi vesivideo. Tähän esitettiin osallistuneet koehenkilöt ($n=5$) olivat 23–27 ($\mu=24,6$; $\sigma=1,52$) vuotiaita miehiä. Heistä kolme oli Mediatekniikan laitoksen henkilökuntaa, ja loput ilmoittivat ammatikseen opiskelun.

Esitestin I tulokset on esitetty kuvassa 5.6. Tulosten perusteella havaittu laatu riippuu vääristymäversiosta. Näiden laatutasojen soveltaminen varsinaisiin testeihin ei ollut kuitenkaan mahdollista, koska kolme eniten huononnettua versiota olivat suurista hajonnoista johtuen lähes jokaisessa videokontentissa havaitun laadun suhteen poikkeamattomia. Vain vähiten huononnettu versio erosi muista versioista merkittävästi.



Kuva 5.6: Esitestin I kontenttikohtaiset tulokset. Kontenttien neljä versiota on esitetty generoidun vääristymän voimakkuuden mukaisessa järjestyksessä.

Esitestissä I kolmen eniten huononnetun version laatuarvot vaihtelivat merkittävästi kontenteittain, joten tavoitteen mukaista kahta samaa havait-

tua laatutasoa kaikille kontenteille ei löytynyt. Karkeasti katsottuna videonlaatuarviot jakautuivat ainoastaan kahdelle tasolle jokaisessa videossa. Tästä syystä tehtiin esitesti II, missä videonlaatua kontrolloitaisiin pakettien tiputtamisen sijaan videon sumentamisella.

5.3.2 Esitesti II

Esitestissä II yritettiin löytää videokontentteihin toisistaan eroavia laatu-
tasoja sumentamalla niitä. Tällöin vääristymän voimakkuutta pystyttiin
säättämään lähes portaattomasti.

Esitestissä II ärsykkeenä olivat 16 kuvattua videota, joista lippuvideo toimi
esimerkkivideona. Kaikkien sekvenssien pituudeksi leikattiin 10 sekuntia,
ja jokaisesta sekvenssistä tehtiin viisi eri versiota, joissa sumennuksen määrä
vaihteli. Jokaisen videon kahden hyvän version laadut asetettiin visuaalises-
ti tarkasteltuna melko lähelle toisiaan. Tavoite oli, että niistä toisen version
arvio olisi yhtä suuri muiden videoiden hyvän version kanssa. Samoin toimit-
tiin myös huonojen versioiden osalla. Sen sijaan yhden version laatu pyrittiin
asettamaan hyvien ja huonojen versioiden väliin, jotta arvioinnit eivät asetu
vain kahdelle kohdalle arviointiasteikossa.

Videoita sumennettiin Matlab:lla toteutetulla ohjelmalla. Siinä aluksi luo-
ttiin fspecial-funktiolla Gaussian-tyyppinen neliömäinen suodin, jonka kes-
kihajonta pysyi aina vakiona, mutta suotimen koko muuttui versioittain.
Tämän jälkeen suotimella suodatettiin jokaisen videon jokainen kuva imfilter-
funktiolla. Lopullinen MPEG-4 pakattu video muodostettiin VideoWriter- ja
writeVideo-funktioiden avulla. Jokaista kuvaa version sisällä sumennettiin
yhtä paljon. Pilvivideota ei valittu testiin, koska siihen generoitu sumentu-
neisuus aiheutti häiritsevää nykimistä, jota muissa videoissa ei esiintynyt.
Tämä johtui mahdollisesti siitä, että kyseisessä videossa oli hyvin vähän yk-
sityiskohtia sumennuksen jälkeen, jolloin videonpakkausalgoritmi ei toiminut
enää optimaalisella tavalla.

Esitestin II ensimmäinen osa aloitettiin koehenkilöillä ($n=5$), jotka olivat
laitoksen työntekijöitä. Videokontentteihin oli generoitu sumentuneisuutta
oman visuaalisen arvion mukaan. Näiltä koehenkilöiltä saadun datan mukaan
hyvän ja huonon laadun tasoa vielä hienosäädettiin videoittain. Taulukossa
5.1 on nähtävissä lopulliset esitestissä käytettyjen Gaussian-suotimien koot.

Tämän jälkeen koetta jatkettiin toisella osalla, johon osallistui naiiveja
koehenkilöitä ($n=13$). He olivat 21–30 ($\mu=25,1$; $\sigma=2,69$) vuotiaita miehiä ja
suurin osa heistä oli opiskelijoita. Sukupuolijakaumalla ei arvioitu olevan vai-
kutusta arvioinneissa, koska videokontentit ovat monotonisia ja eivät herätä
tunteita kummassakaan sukupuolella. Kaikilla oli normaali tai normaaliksi
korjattu näkö, paitsi yhdellä koehenkilöllä lähinäöntarkkuuden vertailuarvo

Taulukko 5.1: Esitestin II videoiden versioissa käytetyt suodinkoot.

Video	Suotimen leveys (px)				
Vaahtera	5	6	8	11	12
Kaisla	4	5	6	8	9
Lehti	5	6	8	14	15
Hedelmä	6	7	11	21	23
Vesi	5	15	25	35	220
Rubik	7	9	14	28	31
Kävely	4	6	12	18	21
Lukemis	5	6	8	16	17
Auto	8	9	14	22	29
Hyppynaru	4	5	10	19	21
Hanhi	4	5	7	12	13
Kynttilä	5	6	10	15	16
Pallo	7	8	14	28	33
Kangas	7	8	10	17	20
Maisema	4	5	7	11	13

oli yhtä tasoa heikompana kuin vielä normaalina näöntarkkuutena pidettävä arvo. Tämän koehenkilön arvioinneissa ei kuitenkaan havaittu poikkeavuutta, joten myös hänen arviot otettiin mukaan tuloksiin.

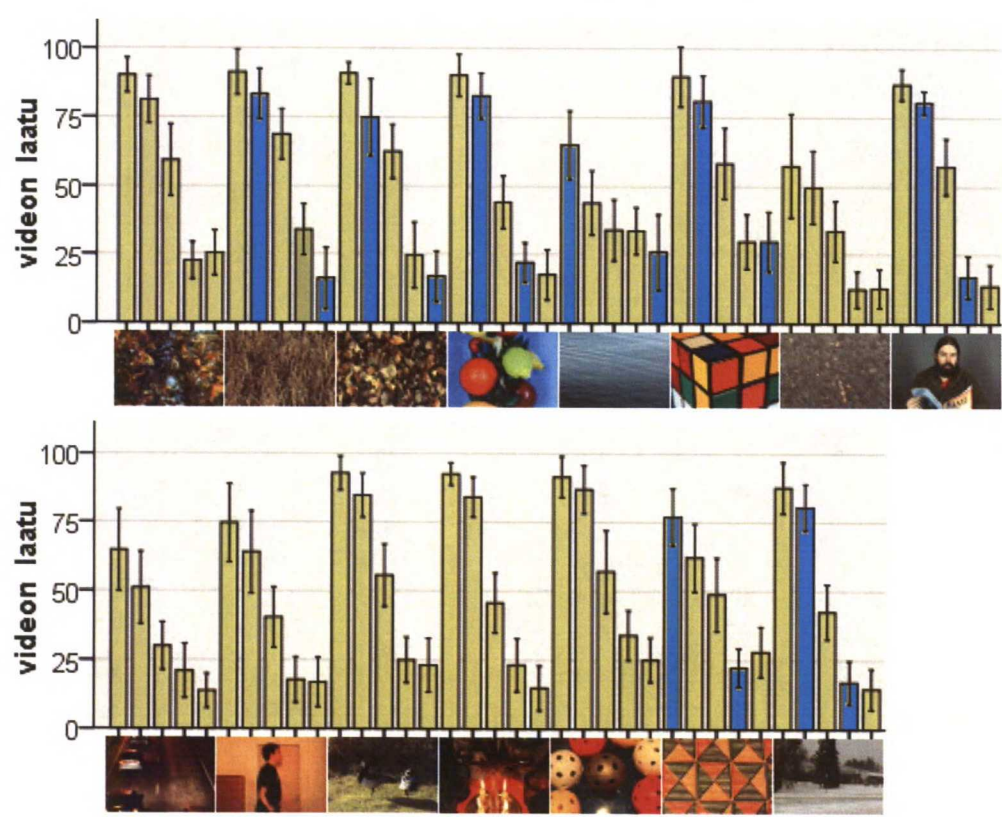
Esitestin lopuksi loppukyselyssä selvitettiin koehenkilöiden mielipiteitä testissä näkemistään videoista. Lisäksi heidän tehtävänään oli mainita, jos joku testivideoista oli heidän mielestään erikoinen, outo, epäluonnollinen, tunteita herättävä tai merkittävästi muista videoista poikkeava. Tämä tehtiin, koska videoita ei ole aikaisemmin käytetty ärsykkeenä missään testissä. Tarkoituksena oli kartoittaa sitä, eroavatko jotkut videokontentit merkittävästi muista videokontenteista jossain asiassa, ja näin löytää mahdollisia syitä, jos tuloksissa on jotain poikkeuksellista. Lisäksi testin jälkeen kysyttiin, mihin koehenkilöiden arviot perustuivat. Koehenkilöt saivat myös antaa vapaamuotoista palautetta testistä.

Erityisesti testin jälkeen saivat kommentteja rubik-, vesi-, ja hyppynaruvideo. Rubik-video sai enimmäkseen positiivisia kommentteja. Sitä keuhuttiin muun muassa ammattimaisesti kuvatuksi. Sen sijaan vesi- ja hyppynaruvideoista koehenkilöt eivät juurikaan pitäneet. Hyppynaruvideossa häiritsi outo valaistus, ja vesivideota monet kuvailivat epäluonnolliseksi. Muiden videoiden kommentteissa oli sen sijaan paljon hajontaa. Esimerkiksi eräät koehenkilöt pitivät paljon kaislavidеosta sen rauhallisuuden takia, kun toisia puolestaan ärsytti videon huono laatu.

Yleisesti videoista mainittiin, että ne olivat koehenkilöiden sanoin mie-

lenkiintoisia, vaihtelevia, monipuolisia ja hyviä. Loppukyselyssä mainittiin, että enemmän ihmisvideoita olisi kaivattu testiin. Ehkä liikkuvien ihmisten tai ihmisruuhkan sisältävä video voisi rikastaa videopankkia.

Varsinaisiin testeihin valittiin kahdeksan kontenttia, jotta testien kesto ei olisi liian pitkä. Varsinaisen testien kontenttivalinta perustui esitestin II tuloksiin. Kuvan 5.7 perusteella kaikista videoista löytyi vähintään yksi hyvää laatua edustava versio noin 75 arvon kohdalta ja vastaavasti yksi huonoa laatua edustava versio noin arvon 20 kohdalta. Näin testeihin oli mahdollista sisällyttää yksi video SI-CI-TI-avaruuden jokaisesta ääripäästä.



Kuva 5.7: Esitestin II videokohdaiset tulokset. Sinisellä värjättyt versiot valittiin varsinaiseen testiin.

Videovalinnassa oli usein valinnanvaraa, koska usein kahdella videokontentilla oli lähes yhtä suuret mitta-arvot. Tästä poikkeuksena olivat kaisla- ja kangasvideot, jotka molemmat valittiin testiin. Lukemisvideo oli itsestään selvä valinta, koska sen kilpailija, vähän liikettä, vähän yksityiskohtia ja vähän värejä sisältävässä sarjassa, oli ainoastaan pilvivideo, joka ei tässä

esitestissä ollut mukana. Lisäksi läheltä henkilöä kuvattu video oli ainut laatuaan.

Kävely- ja vesivideoissa oli paljon liikettä, mutta ei yksityiskohtia tai värejä. Näistä vesivideo valittiin testeihin I ja II, koska vesivideossa oli kävelyvideota lähempänä ääripäitä olevat mitta-arvot. Vesivideossa tosin hyvänlaatuisten versioiden keskiarvo poikkesi hiukan muiden videoiden tasosta. Paljon värejä sisältävistä kynttilä- ja hedelmävideoista valittiin testiin hedelmävideo, koska siinä oli muista videoista poikkeava erityinen pystysuuntainen kameran liike.

Paljon yksityiskohtia sisältävistä hanhi- ja maisemavideoista valittiin testiin maisemavideo, koska samalta etäisyydeltä kuvattuja eläviä objekteja kuin hanhivideossa oli jo testiin valitussa lukemisvideossa. Lisäksi maisemavideo oli ainut video, jossa on rakennuksia ja kaukana kulkevia ihmisiä. Pallo- ja rubik-videoissa oli vähän yksityiskohtia, mutta paljon värejä ja liikettä. Näistä valittiin ehkä hiukan taiteellisesti makro-objektiivilla kuvattu rubik-video. Paljon niin yksityiskohtia, kuin värejä ja liikettä sisältävistä vaahtera- ja lehtivideoista valittiin lehtivideo, koska yksi käsivaralta kuvattu video haluttiin testiin. Esitestissä II oli varavaihtoehtoina myös auto- ja hyppynaruvideo, mutta näitä ei tarvinnut ottaa varsinaisiin testeihin.

Varsinaisiin testeihin valitut versiot eri kontenteista on värjätty sinisellä kuvassa 5.7. Valittujen hyvien ($F(7,96)=1,719$; $p=0,114$) tai huonojen ($F(7,96)=1,116$; $p=0,360$) versioiden arvioissa videokontenttien välillä ei varianssianalyysin perusteella ollut tilastollisesti merkitsevää eroa.

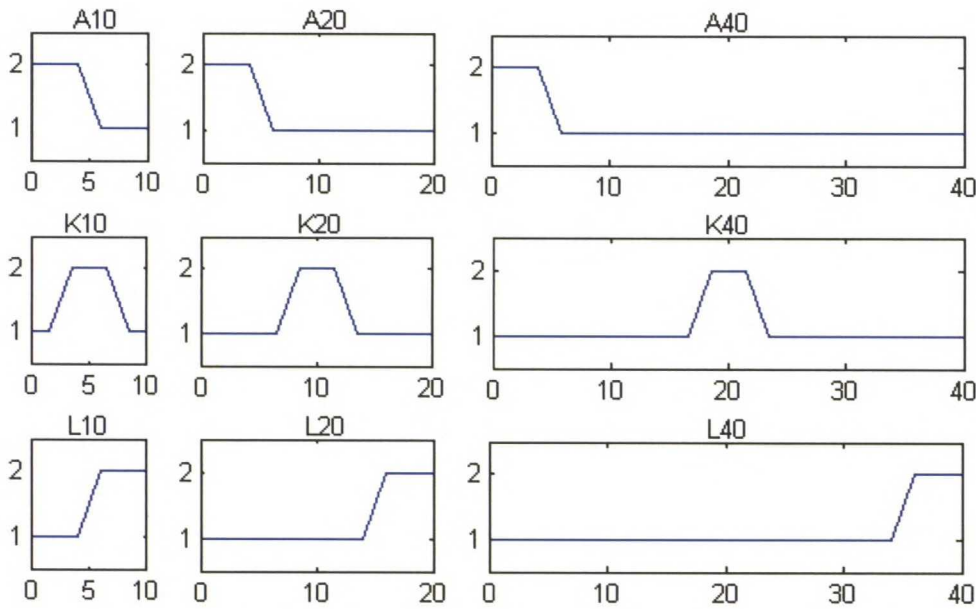
Koehenkilöitä pyydettiin esitesteissä myös listaamaan kokeen päätyttyä niin monta testivideota kuin he muistivat. Tämän tulokset on esitetty liitteessä A.1. Jos jonkun videon arviointitulokset olisivat olleet selvästi muista poikkeavia, olisi videoiden muistettavuudella saattanut pystyä perustelemaan tuloksia. Ihmisiä ja tutun paikan sisältävä maisemavideo oli muistettu parhaiten. Myös eläviä kohteita sisältävät, lukemis- ja hyppynaruvideo olivat hyvin muistettuja. Seesteiset kaisla-, pilvi- ja kangasvideo muistettiin kaikista heikoiten. Tulokset vastaavat melko hyvin kirjallisuudessa [19] esitettyjä tuloksia kuvien muistettavuudesta.

5.4 Testi I: vakiopituinen vääristymä

Testin I tavoite oli selvittää, miten temporaaliset laatumuutokset videosekvenssissä vaikuttavat havaittuun laatuun. Erityisesti oli tarkoitus selvittää, miten videon pituus vaikuttaa tähän. Toisena muuttuvana parametrina toimi huonon laadun sijainti videosekvenssissä. Referenssisisältöjä lukuun ottamatta videoissa esiintyi aina 5 sekuntia huonoa laatua lopun ollessa hyvää

laatua.

Testissä I oli ärsykkeenä esitettiin II perusteella valitut kahdeksan videokontenttia, joista jokaisesta oli arvioitavana 11 erilaista versiota. Yhdeksässä versioista vaihteli videon pituus: 10, 20 tai 40 sekuntia ja se, missä kohtaa sekvenssiä 5 sekunnin sumentunut, huonolaatuinen kohta sijaitsi: alussa (A), keskellä (K) tai lopussa (L) (kuva 5.8). Lisäksi kahdessa versiossa videon koko pituus oli hyvää tai huonoa laatua, ja nämä versiot toimivatkin referenssisisältöinä (REF). Esimerkki version nimeämisestä on K20, missä K kuvaa huonon laadun sijaintia sekvenssissä ja 20 videon pituutta sekunneissa.



Kuva 5.8: Testin I videosekvenssit: A10, A20, A40, K10, K20, K40, L10, L20 ja L40. Vaaka-akselilla kulkee aika sekunneissa. Pystyakselilla on lisätyn vääristymän suuruus.

Luvussa 3.3 esitetyn tutkimuksen [35] mukaan jos vääristymät ovat pidemmällä kuin 15 sekunnin päässä videosekvenssin lopusta, niiden sijainti ei enää tee eroa kokonaislaatuarvioihin. Noin 20–30 sekunnin päässä sekvenssin lopusta sijaitsevien vääristymien pitäisi vaikuttaa enää vähän tai ei ollenkaan videon kokonaislaatuun [4]. Näin 10, 20 ja 40 sekunnin sekvensseillä pystyttiin testaamaan sitä, vastaavatko tämän tutkimuksen tulokset kirjallisuudessa esitettyjä havaintoja. Koe toteutettiin systemaattisesti, koska SI-, TI- ja CI-arvot pysyivät vakioina videokontenteittain, vaikka sekvenssin pituus muuttui. Tämän tutkimuksen tuloksista selviää myös, voiko niihin soveltaa luvussa 3.3 esitettyjä temporaalisen videon laadun painofunktioita [5, 67].

Koehenkilöillä oli arvioitavana yhteensä 88 videota. Versioiden huonon ja hyvän laadun vääristymätasot oli säädetty kontenttikohtaisesti niin, että havaitut huonot ja hyvät laadut olivat samansuuruisia.

Videoihin luodut hyvät ja huonot laatutasot saatiin esitestistä II. Laatu-muutoksia ei luotu videoissa äkkinäisiksi, vaan muutokset tapahtuivat portaattomasti kahden sekunnin aikana. Näin laatumuutoskohtaa ei havaita niin selkeänä. Tällä pyrittiin estämään sitä, että koehenkilöt eivät systemaattisesti arvioisi videonlaatua esimerkiksi sekunteja laskemalla. Hitaahkot laatu-muutokset ovat myös katsojalle miellyttävämpiä kuin nopeat, jolloin laatu-arvio mahdollisesti saataisiin painottumaan enemmän videon kokonaislaatuun kuin laatumuutoskohtiin. Videoiden sumentaminen ja luominen toteutettiin Matlab:lla vastaavasti kuten esitestissä II.

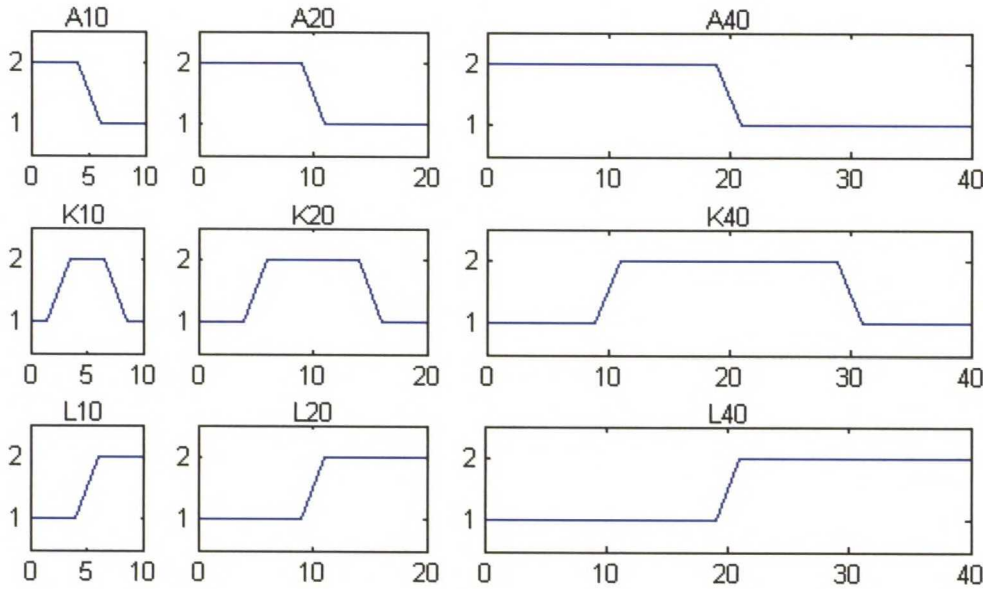
Kokeeseen osallistuneet koehenkilöt ($n=11$) olivat 22–31 ($\mu=25,3$; $\sigma=2,76$) vuotiaita suurimmalta osin opiskelijoita. Heistä kaksi oli naisia. Kaikilla kokeeseen osallistuneilla henkilöillä oli normaali näkö, paitsi yhdellä koehenkilöllä lähinäöntarkkuden vertailuarvo osui vielä normaalina lähinäöntarkkuutta pidettävää tasoa yhtä alemmaksi. Kyseisen koehenkilön arvioinneissa ei kuitenkaan havaittu poikkeavuutta, joten myös hänen arviot otettiin mukaan tuloksiin. Testiin ei osallistunut esitesteissä käyneitä koehenkilöitä, vaan kaikki koehenkilöt näkivät videot ensimmäistä kertaa. Keskimääräinen testin kesto oli noin 55 min.

Testissä I käytettiin luvussa 5.2 kuvattua testimenetelmää. Esimerkki-videona toimi lippuvideo. Testin jälkeisessä loppukyselyssä ei enää kerätty koehenkilöiden mielipiteitä testissä esitetyistä videoista. Sen sijaan siinä kartoitettiin sitä, miten koehenkilöt olivat arvioineet videon, jos he huomasivat, että sen laatu muuttui videon aikana. Tästä nähtiin, jos arvioissa esiintyi yksilöllisiä eroja. Koehenkilöt saivat myös halutessaan antaa vapaamuotoista palautetta testistä.

5.5 Testi II: suhteellinen vääristymä

Testin I tuloksien (luku 6.1) pohjalta heräsi kysymys, miksi 40 sekunnin versioissa lopussa ollut vääristymä ei saanut huonompia arvosanoja kuin alussa tai keskellä ollut vääristymä (ks. kuva 5.8). Sama tilanne ilmeni 20 sekunnin sekvensseissä. Tämä poikkeaa Handsin ja Avonsin [15] sekä Liun ym. [35] tutkimuksien tuloksista. Sen sijaan 10 sekunnin videoissa lopussa ollut vääristymä sai huonompia arvioita kuin alussa tai keskellä ollut vääristymä. Tämä voi johtua joko siitä, että videon pituus oli lyhyt tai että videossa oli paljon huonoa laatua (50 %) suhteessa sen pituuteen. Tähän kysymykseen oli tavoite saada vastaus testistä II, jossa kaikkiin testivideoihin luotiin pro-

sentuaalisesti yhtä paljon huonoa laatua, eli puolet koko videon pituudesta. Testissä II oli ärsykkeenä esitestin II perustella valitut kahdeksan videota, joista jokaisesta oli arvioitavana 11 erilaista versiota. Näin koehenkilöillä oli arvioitavana yhteensä 88 videota. Videosekvenssit on esitetty kuvassa 5.9. Versioiden nimeämisessä käytettiin samaa käytäntöä kuin testissä I.



Kuva 5.9: Testin II videosekvenssit: A10, A20, A40, K10, K20, K40, L10, L20 ja L40. Vaaka-akselilla kulkee aika sekunneissa. Pystyakselilla on lisätyn vääristymän suuruus.

Videoihin luodut hyvät ja huonot laatutasot saatiin esitestistä II. Laatu-muutokset tapahtuivat portaattomasti testin I tapaan. Videoiden sumenta-minen ja luominen toteutettiin Matlab:lla vastaavasti kuten esitestissä II ja testissä I. Testissä II käytettiin luvussa 5.2 kuvattua testimenetelmää. Esi-merkkivideonä toimi lippuvideo. Testin jälkeisessä loppukyselyssä kysyttiin sitä, miten koehenkilöt olivat arvioineet videon, jos he huomasivat, että sen laatu muuttui videon aikana.

Testiin II osallistuneet koehenkilöt ($n=12$) olivat 21–44 ($\mu=28,3$; $\sigma=5,59$) vuotiaita, joista 2 oli naisia. He olivat pääosin opiskelijoita. Koehenkilöillä oli normaali tai normaaliksi korjattu lähinäkö ja normaali värinäkö. Kaikki koehenkilöt näkivät videot ensimmäistä kertaa. Testin keskimääräinen kesto oli taas noin 55 minuuttia.

Luku 6

Tulokset ja analyysi

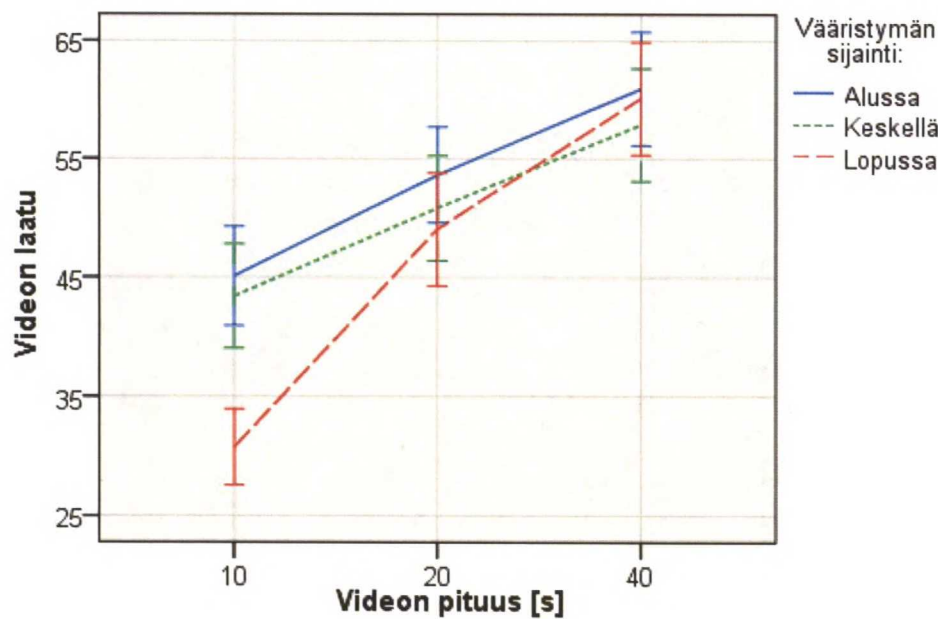
Tässä luvussa on esitetty testien I ja II tulokset ja analyysi.

6.1 Testi I: vakio pituinen vääristymä

Testin I tuloksiin ei ole sisällytetty koehenkilö 11:n arviointeja. Koehenkilö 11 ei ollut ymmärtänyt ohjeistusta ja oli sijoittanut kaikki videonlaatu-arviot kapealle alueelle (10–71) asteikon keskivaiheelle. Tämä on nähtävissä koehenkilöiden arviojakautumista esittämässä liitteestä B.3. Muiden koehenkilöiden arviointikäyttäytymisessä ei havaittu poikkeavuutta.

Kuvassa 6.1 on esitetty testin I keskimääräiset videoversioiden laatu-arviot 95 %:n luottamusväliä käyttäen. Versioissa videon pituus ja 5 sekunnin vääristymän sijainti ovat olleet muuttujina. Yksisuuntaisen varianssianalyysin perusteella joidenkin versioiden välillä on tilastollisesti merkitsevää eroa ($F(10,869)=46,932, p=0,000$). Versioiden välisiä eroja tutkittiin Tukeyn Post Hoc -testin avulla. Taulukossa 6.1 on nähtävillä kahdeksan osajoukkoa, joihin kuuluvien versioiden laatu-arvioiden keskiarvot eivät poikke tilastollisesti toisistaan. Siitä on nähtävissä, että 40 sekunnin tai 20 sekunnin pituisien sekvenssien eri versioiden välillä ei ole tilastollisesti merkitsevää eroa. Sen sijaan 10 sekunnin sekvenssin version L10 laatu-arvio (30,78) eroaa versioista A10 (45,10, $p=0,000$) ja K10 (43,41, $p=0,003$) tilastollisesti merkitsevästi. Molemmat referenssivideot eroavat tilastollisesti merkitsevästi kaikista muista versioista, mikä vielä vahvistaa sen, että vääristymät ovat aina olleet havaittavissa ja vaikuttaneet laatu-arvioissa.

Vääristymän sijainnilla videosekvenssissä ei tämän testin perusteella ollut vaikutusta havaitussa laadussa, jos videon pituus oli 20 tai 40 sekuntia. Sen sijaan jos videon pituus oli 10 sekuntia, oli videon laatu havaittu heikompana, jos vääristymä oli videon lopussa kuin jos se olisi ollut videon alussa tai



Kuva 6.1: Viisi sekuntia kestävän vääristymän sijainnin ja videosekvenssin pituuden vaikutus havaittuun laatuun (95 %:n luottamusvälit).

Taulukko 6.1: Testin I kahdeksan osajoukkoa, joiden sisällä versioiden havaitulla laadulla ei ole tilastollisesti merkitsevää eroa. Videon pituus on 10, 20 tai 40 sekuntia, ja vääristymä on videon alussa (A), keskellä (K) tai lopussa (L).

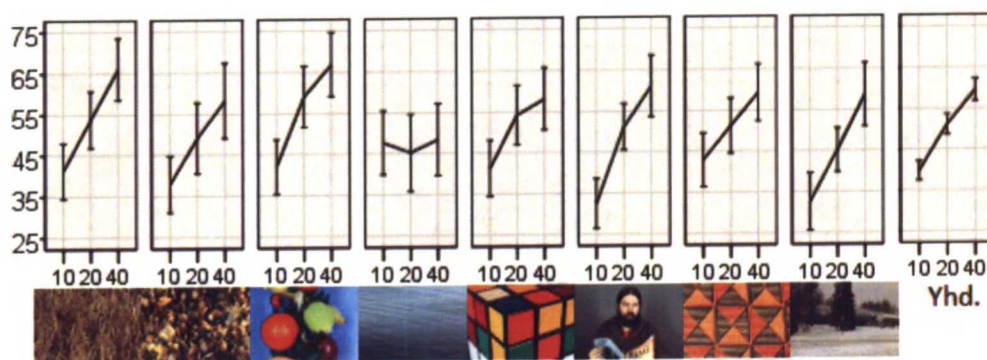
Versio	Alfa osajoukolle = 0,05							
	1	2	3	4	5	6	7	8
REF10	18,79							
L10		30,78						
K10			43,41					
A10			45,10	45,10				
L20			49,04	49,04	49,04			
K20			50,81	50,81	50,81	50,81		
A20				53,61	53,61	53,61	53,61	
K40					57,86	57,86	57,86	
L40						60,10	60,10	
A40							60,91	
REF10								73,81
p	1,000	1,000	,382	,186	,146	,100	,404	1,000

keskellä. Tämä 10 sekunnin tilanne voi johtua siitä, että video oli lyhyt tai siitä että videon vääristymän pituus oli prosentuaalisesti suurempi kuin 20 tai 40 sekunnin tilanteissa.

Luvussa 2.2.5 esitettyä muistiefektiä on havaittavissa ainoastaan 10 sekunnin videosekvensseissä. Tällöin videon lopussa ollut vääristymä on laske-
nut videon laatuarviota enemmän kuin kauempana videon loppua sijainneet vääristymät. 10 sekunnin videosekvenssien tulokset vastaavat Zinkin ym. [68] tutkimuksessa saatuja tuloksia. Sen sijaan 20 ja 40 sekunnin videosekvenssien tulokset eivät vastaa aiemmissa tutkimuksissa esitettyjä tuloksia, koska niissä vääristymän sijainti ei ole vaikuttanut laatuarvioon.

Kun tarkastellaan kaikkia 10, 20 ja 40 sekunnin versioita yhdessä, huomataan, että videon pituudella on tilastollisesti merkitsevä vaikutus havaitussa videon laadussa ($F(2,717)=59,317$; $p=0,000$). Tukeyn Post Hoc -testin perusteella 10 sekunnin (39,76) videoiden laatu eroaa tilastollisesti merkitsevästi 20 sekunnin videoista (51,15; $p=0,000$) ja 20 sekunnin videoiden laatu eroaa tilastollisesti merkitsevästi 40 sekunnin videoista (59,63; $p=0,000$). Tämä tulos on erittäin looginen, koska mitä lyhyempi video on, sitä enemmän siinä on vääristynyttä laatua prosentuaalisesti. 10 sekunnin videoissa vääristymää oli puolet koko videon pituudesta, kun puolestaan 20 sekunnin tilanteessa vääristymää oli neljännesosa ja 40 sekunnin tilanteessa vain kahdeksasosa koko videon pituudesta.

Koska tulos videon pituuden vaikutuksesta havaitussa laadussa oli selkeä, voidaan tarkastella sitä, miten videokontentti on vaikuttanut arviointeihin. Kuvassa 6.2 on esitetty videokontentin vaikutus videon laatuarvioon, kun videon pituus on muuttunut.



Kuva 6.2: Videokontentin vaikutus havaittuun laatuun videon pituuden (10s/20s/40s) muuttuessa.

Kaikkien videokontenttien tuloksissa on havaittavissa selkeä nouseva

käyrä, mutta vesivideon tilanteessa arvot ovat pysyneet samalla tasolla kaikilla videon pituuksilla. Vesivideon lisäksi rubik- ja kangasvideoista ei löytynyt tilastollisesti merkitsevää vaikutusta, kun videon pituus muuttuu. Vesivideo on kuitenkin joukon ainoa, jossa arvot poikkeavat muista videokontenteista näin paljon. Tämä saattaa johtua siitä, että vesivideossa on vähän yksityiskohtia ja värejä, jolloin videon sumentaminen ei juurikaan ole vaikuttanut havaittuun laatuun.

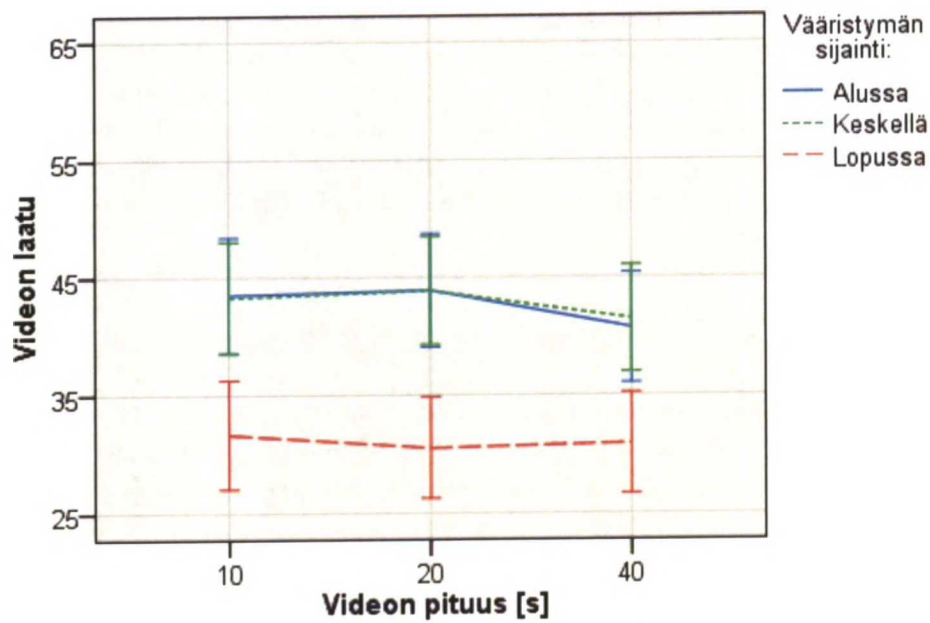
6.2 Testi II: suhteellinen vääristymä

Testin II koehenkilöiden arviojakaumat on nähtävissä liitteessä B.3. Koehenkilö 7 oli sijoittanut kaikki videonlaatuarviot pienelle alueelle (29–56) asteikon keskivaiheelle. Koehenkilö 9 oli käyttänyt vain arviointiasteikon alapäätä (6–69). Tästä johtuen koehenkilöt 7 ja 9 rajattiin jatkokäsittelyn ulkopuolelle.

Kuvassa 6.3 on esitetty testin II keskimääräiset videoversioiden laatuarviot 95 %:n luottamusväliä käyttäen. Versioissa videon pituus ja vääristymän sijainti ovat olleet muuttujina. Vääristymän pituus oli kaikissa sekvensseissä puolet sekvenssin pituudesta. Yksisuuntaisen varianssianalyysin perusteella eri versioiden välillä on tilastollisesti merkitsevää eroa ($F(10,869)=44,975$, $p=0,000$). Versioiden välisiä eroja tutkittiin Tukeyn Post Hoc -testin avulla. Taulukossa 6.2 on nähtävillä versiot, joiden keskiarvot poikkeavat tilastollisesti toisistaan. Tukeyn Post Hoc -testin perusteella versio L10 (31,69) eroaa tilastollisesti merkitsevästi versiosta A10 (43,51, $p=0,000$) ja K10 (43,29, $p=0,000$). Samoin versio L20 (30,55) eroaa tilastollisesti merkitsevästi versiosta A20 (43,90, $p=0,002$) ja K20 (43,86, $p=0,002$). Sen sijaan versio L40 (30,93) ei eroa tilastollisesti merkitsevästi versiosta A40 (40,70, $p=0,098$) tai K40 (41,46, $p=0,050$), vaikka myös L40 version keskiarvo on pienempi kuin versioiden A40 ja K40 keskiarvo.

Tuloksista on nähtävissä, että videon tai vääristymän suhteellinen pituus ei ole tehnyt eroa havaittuun videon laatuun. Tässä on havaittavissa luvussa 2.2.5 esitettyä pituuden laiminlyöntiä, jossa vääristymän tai videon pituus ei vaikuta arviointeihin. Vääristymän sijainti on sen sijaan vaikuttanut havaittuun laatuun. Trendi on jokaisella videon pituudella sama: videon alussa ja keskellä olleet vääristymät ovat tuottaneet suurempia videon laatuarvioita kuin videon lopussa sijainneet vääristymät.

Kun 10, 20 ja 40 sekunnin pituisia videoita vertaillaan yksisuuntaisen varianssianalyysin avulla huomataan, että videon pituudella ei ole ollut tilastollisesti merkitsevää vaikutusta havaittuun laatuun ($F(2,717)=0,549$; $p=0,578$). Sen sijaan kun verrataan A, K ja L versioita keskenään huoma-



Kuva 6.3: Puolet videon pituudesta kestävä vääristymän sijainnin ja videosekvenssin pituuden vaikutus havaittuun laatuun (95 %:n luottamusvälit).

Taulukko 6.2: Testin II viisi osajoukkoa, joiden sisällä versioiden havaitulla laadulla ei ole tilastollisesti merkitsevää eroa. Videon pituus on 10, 20 tai 40 sekuntia, ja vääristymä on videon alussa (A), keskellä (K) tai lopussa (L).

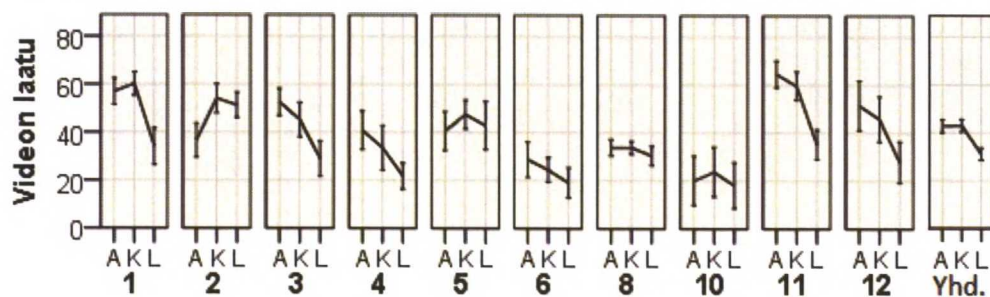
Versio	Alfa osajoukolle = 0,05				
	1	2	3	4	5
REF10	14,95				
L20		30,55			
L40		30,93	30,93		
L10		31,69	31,69		
A40		40,70	40,70	40,70	
K40			41,46	41,46	
K10				43,29	
A10				43,51	
K20				43,86	
A20				43,90	
REF10					78,45
p	1,000	,071	,050	,996	1,000

taan, että vääristymän sijainnilla videossa on ollut tilastollisesti merkitsevä vaikutus havaittuun laatuun ($F(2,717)=25,787$; $p=0,000$). Kun tarkastellaan tilannetta Tukeyn Post Hoc -testillä, on todettavissa, että L versio (31,05) eroaa tilastollisesti merkitsevästi versioista A (42,70; $p=0,000$) ja K (42,87; $p=0,000$). Versiot A ja K eivät eroa tilastollisesti merkitsevästi toisistaan ($p=0,996$).

Kymmenen sekunnin tulokset vastaavat testin I samojen versioiden arvioita. Muistiefektiä voidaan havaita kaikilla eripituisilla videoilla, koska lopussa ollut vääristymä on aina arvioitu heikoimmilla arvosanoilla kuin alussa tai keskellä sijainnut vääristymä. Tulosten mukaan muistiefektin kestolla ei ole yleistettävää pituutta, koska 10, 20 ja 40 sekunnin videosekvenssit saavat jokaisessa tilanteessa samat laatuarviot. Tulosten perusteella ainakin lyhyillä videoilla (<40 s) muistiefektin kesto on suhteellinen, eikä absoluuttinen ajan suhteen.

Videokontentti ei vaikuttanut tilastollisesti merkitsevästi ($F(23,696)=0,642$, $p=0,900$) siihen, miten eripituiset videot arvioitiin. Videon pituudella ei myöskään ollut kenenkään koehenkilön kohdalla tilastollista merkitystä laadun arvioinnissa.

Koehenkilökohtaisia arvioita on verrattu vääristymän sijaintiin kuvassa 6.4. Tukeyn Post Hoc -perusteella vääristymän sijainnin vaihtuessa, oli puolella koehenkilöistä tilastollisesti merkitsevä ero havaitussa videon laadussa. Koehenkilöt 1 ja 11 arvioivat videon, jonka lopussa on vääristymä, tilastollisesti merkitsevästi heikompana kuin videon, jonka alussa tai keskellä on vääristymä. Koehenkilöt 3, 4 ja 12 arvioivat videon, jonka lopussa on vääristymä, tilastollisesti merkitsevästi heikompana kuin videon, jonka alussa on vääristymä. Lopuilla viidellä koehenkilöllä vääristymän sijainnilla sekvenssissä ei todettu olevan tilastollista merkitystä videon laatuarviossa.



Kuva 6.4: Koehenkilökohtaiset arviot videosekvenssin alkuun (A), keskelle (K) tai loppuun (L) sijoittuville vääristymille.

Koehenkilöillä on nähtävillä pieniä yksilöllisiä eroja siinä, miten vääristymän

sijainti videossa on vaikuttanut havaittuun laatuun. Useimmiten videon lopussa ollut vääristymä on arvioitu kaikista häiritsevimpänä, kun taas videon alussa tai keskellä olleet vääristymät on arvioitu yhtä suurina. Tästä poikkeuksena on toinen koehenkilö, joka on antanut kaikista huonoimmat arvosanat videoille, joiden alussa on vääristymä. Vääristymän sijainti ei ole kuitenkaan tilastollisesti merkitsevästi vaikuttanut hänen arviointeihin.

Missään kummankaan testin tilanteessa havaittua videon laatua ei painotettu You ym. [67] esittämän funktion mukaisesti.

Luku 7

Johtopäätökset

Tässä diplomityössä tutkittiin videon laadun temporaalisia laatumuutoksia. Tutkimuksen tavoitteena oli saada selville, miten videosekvenssin pituus ja siinä olevan vääristymän sijainti vaikuttavat havaittuun videon laatuun. Tutkimusta varten kuvattiin videoita, joiden sisältö ei muuttunut merkittävästi videosekvenssien aikana ja sekvenssit olivat tarpeeksi pitkiä, mikä mahdollisti eripituisten sekvenssien vertailun.

Työn kirjallisuusosassa esiteltiin muun muassa videon laatuun vaikuttavia tekijöitä ja videon laadun arvioinnin menetelmiä. Erityisesti painopiste oli videon laadun temporalisten muutoksien havaitsemisessa. Kirjallisuusosassa nousi esiin niin sanottu muistiefekti, jolla on todettu olevan vaikutusta videon laatuarvioissa. Videon loppuosan laatu muistetaan paremmin kuin videon alkuosan laatu, joten videon lopun laatu vaikuttaa muita osia enemmän videolle annettuun kokonaislaatuarvioon.

Kuvattuja videoita, joihin oli simuloitu vääristymiä, käytettiin ärsykkeenä subjektiivisissa testeissä, joissa koehenkilöt arvioivat videoiden kokonaislaatua. Testien tuloksien perusteella vääristymän suhteellinen pituus vaikutti siihen, miten voimakkaasti videosekvenssin loppupäätä painotettiin kokonaislaatuarviossa. Sen sijaan videosekvenssien pituus (10s, 20s tai 40s) ei vaikuttanut laatuarvioon, kun vääristymän pituus säilyi suhteellisesti samana. Jos vääristymän pituus oli puolet videosekvenssin pituudesta, sekvenssin lopussa sijainneella vääristymällä oli alentava vaikutus laatuarvioissa. Jos vääristymän pituus oli neljäsosa tai kahdeksasosa sekvenssin pituudesta, vääristymän sijainti ei enää vaikuttanut videon havaittuun laatuun.

Tuloksista nähdään, että objektiivisten laatumittojen tarkkuutta voidaan parantaa videokuvien temporalisella painotuksella. Merkittävä tutkimuksen tulos oli se, että muistiefekti vaikuttaa tuloksiin vain, jos riittävän suuri osuus videosta on vääristynyttä. Videokuvia temporalisesti painottavaa funktiota ei voi laatia suhteessa absoluuttiseen aikamuuttuun tai videon pituuteen,

vaan se pitää laatia suhteessa vääristymän suhteelliseen pituuteen ja sen sijaintiin.

Kirjallisuudessa esitettiin, että 30 sekunnin videosekvenssin laatuarvioon vaikuttaa se, onko siinä sijaitsevan 5 sekunnin vääristymä sekvenssin alussa vai lopussa. Kirjallisuudessa esitettiin myös, että pidemmällä kuin 15 sekunnin päässä videosekvenssin lopusta sijaitsevat vääristymät eivät enää tee eroa videon kokonaislaatuarvioon. Nämä tulokset eivät vastaa tämän tutkimuksen tuloksia. Näin ollen on mahdollista, että vääristymän voimakkuudella, joka ei ollut tämän työn varsinaisissa testeissä muuttuvana parametrina, on vaikutusta muistiefektin voimakkuuteen. Jatkossa olisi mielenkiintoista tutkia, miten vääristymän voimakkuus vaikuttaa muistiefektin suuruuteen samoja videokontentteja käyttäen.

Tutkimuksessa valmistetun videopankin alkuperäisiä videoita voi käyttää hyödyksi temporaalista videon laatua mittaavissa subjektiivisissa testeissä. Tämän vahvistavat työssä onnistuneesti suoritettut koehenkilötestit ja niistä saadut konsistentit ja merkitsevät tulokset. Videopankkia voi hyödyntää videon laatua arvioivien algoritmien optimoinnissa ja videon prosessointisysteemien tai palveluiden laadun kehittämisessä. Se sopii hyvin käytettäväksi täyden referenssin objektiivisten videon laatumittojen tutkimiseen, koska alkuperäiset referenssivideot ovat saatavilla. Videopankki on vapaasti käytettävissä ja sitä voi tiedustella työn tekijältä.

Lähdeluettelo

- [1] ACHANTAY, R., HEMAMIZ, S., ESTRADAY, F., SUSSTRUNK, S. Frequency-tuned salient region detection. *IEEE Conference on Computer Vision and Pattern Recognition* (2009), s. 1597-1604.
- [2] ADOBE DIGITAL INDEX. *The U.S. Digital Video Benchmark - 2012 Review*. <http://success.adobe.com/en/na/programs/digital-index/1304-13926-online-video-report.html> (Viitattu 19.11.2013).
- [3] AHMAD, J., FAYSAL, M. R. Image-processing pipeline for highest quality images. *World Academy of Science, Engineering and Technology*, Vol. 35, (2009), s. 216.
- [4] ALDRIDGE, R., DAVIDOFF, J., GHANBARI, M., HANDS, D., PEARSON, D. Recency effect in the subjective assessment of digitally-coded television pictures. *Fifth International Conference on Image Processing and its Applications*, (1995), s. 336-339.
- [5] BARKOWSKY, M., ESKOFIER, B., BITTO, R., BIALKOWSKI, J., KAUP, A. Perceptually motivated spatial and temporal integration of pixel based video quality measures. *Proceeding MobConQoE '07, Welcome to Mobile Content Quality of Experience*, No. 4, (2007), s. 1-7.
- [6] BARTEN, P. The effects of picture size and definition on perceived image quality. *IEEE Transactions on Electron Devices*, Vol. 36, No. 9, (1989), s. 1865-1869.
- [7] BRAUN, J. Worldwide tv unaffected by the crisis! *Eurodata TV Worldwide*, 2010.
- [8] CHIKKERUR, S., SUNDARAM, V., REISSLEIN, M., KARAM, L. J. Objective video quality assessment methods: A classification, review, and performance comparison. *IEEE Transactions On Broadcasting*, Vol. 57, No. 2, (2011), p. 165-182.

- [9] EEROLA, T., LENSU, L., KAMARAINEN, J.-K., LEISTI, T., RITALA, R., NYMAN, G., KÄLVIÄINEN, H. Bayesian network model of overall print quality: construction and structural optimisation. *Pattern recognition letters*, Vol. 32, No. 11, (2011), s. 1558-1566.
- [10] ENGELDRUM, P. G. Psychometric scaling: Avoiding the pitfalls and hazards. *Society for Imaging Science and Technology: Image Processing, Image Quality, Image Capture Systems Conference*, (2001), s. 101-107.
- [11] GARDINER, P. N., GHANBARI, M., PEARSON, D. E., TAN, K. T. Development of a perceptual distortion meter for digital video. *International Broadcasting Convention*, (1997), s. 493-497.
- [12] GOLDSTEIN, E. B. *Sensation and Perception (8th Edition)*. Wadsworth-Thomson Learning, 2010.
- [13] HALONEN, R., WESTMAN, S., OITTINEN, P. Naturalness and interestingness of test images for visual quality evaluation. *Proceedings SPIE, Image Quality and System Performance VIII*, Vol. 7867, (2011).
- [14] HANDS, D. S. Temporal characterisation of forgiveness effect. *Electronics Letters*, Vol. 37, No. 12, (2007), s. 752-754.
- [15] HANDS, D. S., AVONS, S. E. Recency and duration neglect in subjective assessment of television picture quality. *Applied Cognitive Psychology*, Vol. 15, No. 6, (2001), s. 639-657.
- [16] HASLERA, D., SUSSTRUNK, S. Measuring colourfulness in natural images. *Proceedings IS&T/SPIE Electronic Imaging 2003: Human Vision and Electronic Imaging VIII*, Vol. 5007, (2003), s. 87-95.
- [17] HEKSTRA, A., BEERENDS, J., LEDERMANN, D., DE CALUWE, F., KOHLER, S., KOENEN, R., RIHS, S., EHRSAM, M., SCHLAUSS, D. Pvqm-a perceptual video quality measure. *Signal Processing: Image Communication*, Vol. 17, No. 10, (2002), s. 781-798.
- [18] INTERNATIONAL IMAGING INDUSTRY ASSOCIATION (I3A). *CPIQ Initiative Phase 1 White Paper: Fundamentals and review of considered test methods*, 2007.
- [19] ISOLA, P., XIAO, J., TORRALBA, A., OLIVA, A. What makes an image memorable? *IEEE Conference on Computer Vision and Pattern Recognition*, (2011), s. 145-152.

- [20] ITU-R BT.1788. *Methodology for the subjective assessment of video quality in multimedia applications*, 2007.
- [21] ITU-R BT.500-13. *Methodology for the subjective assessment of the quality of television pictures*, 2012.
- [22] ITU-T E.800. *Definitions of terms related to quality of service*, 2008.
- [23] ITU-T J.143. *User requirements for objective perceptual video quality measurements in digital cable television*, 2000.
- [24] ITU-T P.10/G.100. *Vocabulary for performance and quality of service. Amendment 1: New Appendix I - Definition of Quality of Experience (QoE)*, 2006.
- [25] ITU-T P.910. *Subjective video quality assessment methods for multimedia applications. Audiovisual quality in multimedia services*, 2008.
- [26] JUMISKO-PYYKKÖ, S. Produced quality is not perceived quality - a qualitative approach to overall audiovisual quality. *3DTV Conference*, (2007), s. 1-4.
- [27] JUMISKO-PYYKKÖ, S., HÄKKINEN, J. Evaluation of subjective video quality of mobile devices. *Proceedings of the 13th annual ACM international conference on Multimedia*, (2005), s. 535-538.
- [28] KEELAN, B. *Handbook of Image Quality: Characterization and Prediction*. Marcel Dekker, 2002.
- [29] KNOBLICH, G., THORNTON, I., GROSJEAN, M., SHIFFRAN, M. *Human Body Perception from the Inside Out*. Oxford University Press, 2005.
- [30] KORHONEN, J., REITER, U. Analysis on the perceptual impact of bit errors in practical video streaming applications. *Proceedings of the 3rd IEEE international conference on Internet multimedia services architecture and applications*, (2009), s. 1-6.
- [31] LEA NUMBERS. *Near Vision Card. #27090*. <http://www.lea-test.fi/index.html?start=su/nakotest/instruct/instruct.html> (Viitattu 19.11.2013).
- [32] LEDERER, S., MULLER, C., TIMMERER, C. Dynamic adaptive streaming over http dataset. *Proceedings of the 3rd Multimedia Systems Conference*, (2012), s. 89-94.

- [33] LEE, J-S., GOLDMANN, L., EBRAHIMI, T. Paired comparison-based subjective quality assessment of stereoscopic images. *Multimedia Tools and Applications*, Vol. 67, No. 1, (2013), 31-48.
- [34] LEONTARIS, A., COSMAN, P., REIBMAN, A. Quality evaluation of motion-compensated edge artifacts in compressed video. *IEEE Transactions on Image Processing*, Vol. 16, No. 4, (2011), s. 943-956.
- [35] LIU, T., WANG, Y., BOYCE, J. M., YANG, H., WU, Z. A novel video quality metric for low bit-rate video considering both coding and packet-loss artifacts. *IEEE Journal Of Selected Topics In Signal Processing*, Vol. 3, No. 2, (2009), s. 280-293.
- [36] MU, M., GOSTNER, R., MAUTHE, A., TYSON, G., GARCIA, F. Visibility of individual packet loss on h.264 encoded video stream - a user study on the impact of packet loss on perceived video quality. *Proceedings SPIE, Multimedia Computing and Networking*, Vol. 7253, (2003).
- [37] MUNSELL COLOR. *The Farnsworth D-15 Test*. http://www.munsell.eu/html/colour_vision_tests.html (Viitattu 19.11.2013).
- [38] MURDOCH, M. J., VAN ETTEN, M., HEYNDERICKX, I. E. J. Effect of environmental factors on perception of image attributes. *ICIS 2010: International Congress of Imaging Science*.
- [39] THE NIELSEN COMPANY. *Viewing on Demand - The Cross-Platform Report, The Cross-Platform Series*, 2013.
- [40] NUUTINEN, M. Reduced-reference methods for measuring quality attributes of natural images in imaging systems. Väitöskirja, Aalto-yliopisto, 2012.
- [41] NYMAN, G., LEISTI, T., LINDROOS, P., RADUN, J., SUOMI, S., VIRTANEN, T., OLIVES, J-L., OJA, T., VUORI, T. Measuring multivariate subjective image quality for still and video cameras and image processing system components. *Image Quality and System Performance V: proceedings of the IS&T/SPIE's International Symposium on Electronic Imaging*, Vol. 6808, (2008).
- [42] PEARSON, D. Viewer response to time-varying video quality. *SPIE: Human Vision and Electronic Imaging III*, Vol. 3299, (1998).
- [43] PEATROSS, J., WARE, M. *Physics of Light and Optics*, 2008. <http://optics.byu.edu/textbook.aspx> (Viitattu 19.11.2013).

- [44] PINSON, M. H., WOLF, S. Comparing subjective video quality testing methodologies. *SPIE Video Communications and Image Processing Conference*, Vol. 5150, (2003), s. 573-582.
- [45] PINSON, M. H., WOLF, S. A new standardized method for objectively measuring video quality. *IEEE Transactions On Broadcasting*, Vol. 50, No. 3, (2004), s. 312-322.
- [46] RADUN, J., LEISTI, T., VIRTANEN, T., HÄKKINEN, J., VUORI, T., NYMAN, G. Evaluating the multivariate visual quality performance of image-processing components. *ACM Transactions on Applied Perception (TAP)*, Vol. 7, No. 3, (2010), s. 1-16.
- [47] RAMANATH, R., SNYDER, W. E., YOO, Y., DREW, M. S. Color image processing pipeline - a general survey of digital still camera processing. *IEEE Signal Processing Magazine*, Vol. 22, No 1., (2005), s. 34-43.
- [48] RICHARDSON, I. E. *The H.264 Advanced Video Compression Standard, Second Edition*. Wiley, 2010.
- [49] RIMAC-DRLJE S., VRANJES M., ZAGAR D. Influence of temporal pooling method on the objective video quality evaluation. *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, 2009.
- [50] ROHALY, A. M., LU, J., FRANZEN, N. R., RAVEL, M. K. Comparison of temporal pooling methods for estimating the quality of complex video sequences. *Proceedings SPIE: Human Vision and Electronic Imaging IV*, Vol. 3644, (1999).
- [51] SEFERIDIS, V., GHANBARI, M., PEARSON, D. E. Forgiveness effect in subjective assessment of packet video. *Electronics Letters*, Vol. 28, No. 21, (1992), s. 2013-2014.
- [52] SEKULER, R., BLAKE, R. *Perception (5th edition)*. McGraw-Hill Higher Education, 2005.
- [53] SESHADRINATHAN, K., BOVIK, A. C. Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Transactions On Image Processing*, Vol. 19, No. 2, (2010), s. 335-350.
- [54] SESHADRINATHAN, K., SOUNDARARAJAN, R., BOVIK, A. C., CORMACK, L. K. Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing*, Vol. 19, No. 6, (2010), s. 1427-1441.

- [55] SESHADRINATHAN, K., SOUNDARARAJAN, R., BOVIK, A. C., CORMACK, L. K. A subjective study to evaluate video quality assessment algorithms. *SPIE Proceedings Human Vision and Electronic Imaging*, Vol. 7527, (2010).
- [56] STOCKHAMMER, T., HANNUKSELA, M., WIEGAND, T. H.264/avc in wireless environments. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, (2003), s. 657-673.
- [57] TAKAHASHI, A., HANDS, D., BARRIAC, V. Standardization activities in the itu for a qoe assessment of iptv. *IEEE Communications Magazine*, Vol. 46, No. 2, (2008), s. 78-84.
- [58] VQEG. *Final Report From The Video Quality Experts Group On The Validation Of Objective Models Of Video Quality Assessment*, 2000.
- [59] VRANJES, M., RIMAC-DRLJE, S., GRGIC, K. Review of objective video quality metrics and performance comparison using different databases. *IEEE Signal Processing: Image Communication*, Vol. 28, No. 1, (2013), s. 1-19.
- [60] WALKER, B. *Optical Engineering Fundamentals*. The International Society for Optical Engineering, 2009.
- [61] WANG, Z., BOVIK, A. C. Mean squared error: Love it or leave it? - a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, Vol. 26, No. 1, (2009), s. 98-117.
- [62] WANG, Z., LU, L., BOVIK, A. C. Video quality assessment based on structural distortion measurement. *IEEE Signal Processing: Image Communication*, Vol. 19, No. 2, (2004), s. 121-132.
- [63] WANG Z., SHANG X. Spatial pooling strategies for perceptual image quality assessment. *IEEE International Conference on Image Processing*, 2006.
- [64] WIKSTRAND, G. Improving user comprehension and entertainment in wireless streaming media - introducing cognitive quality of service. Li-sensiaattityö, Umeå University, 2003.
- [65] WINKLER, S. *Digital Video Quality: Vision Models and Metrics*. John Wiley & Sons, 2005.

- [66] WINKLER, S. Analysis of public image and video databases for quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, Vol. 6, No. 6, (2012), s. 616-625.
- [67] YOU J., KORHONEN J., PERKIS A. Attention modeling for video quality assessment: Balancing global quality and local quality. *IEEE International Conference on Multimedia and Expo (ICME)*, 2010.
- [68] ZINK, M., KÖNZEL, O., SCHMITT, J., STEINMETZ, R. Subjective impression of variations in layer encoded videos. *IWQoS'03 Proceedings of the 11th international conference on Quality of service*, (2003), s. 137-154.

Liite A

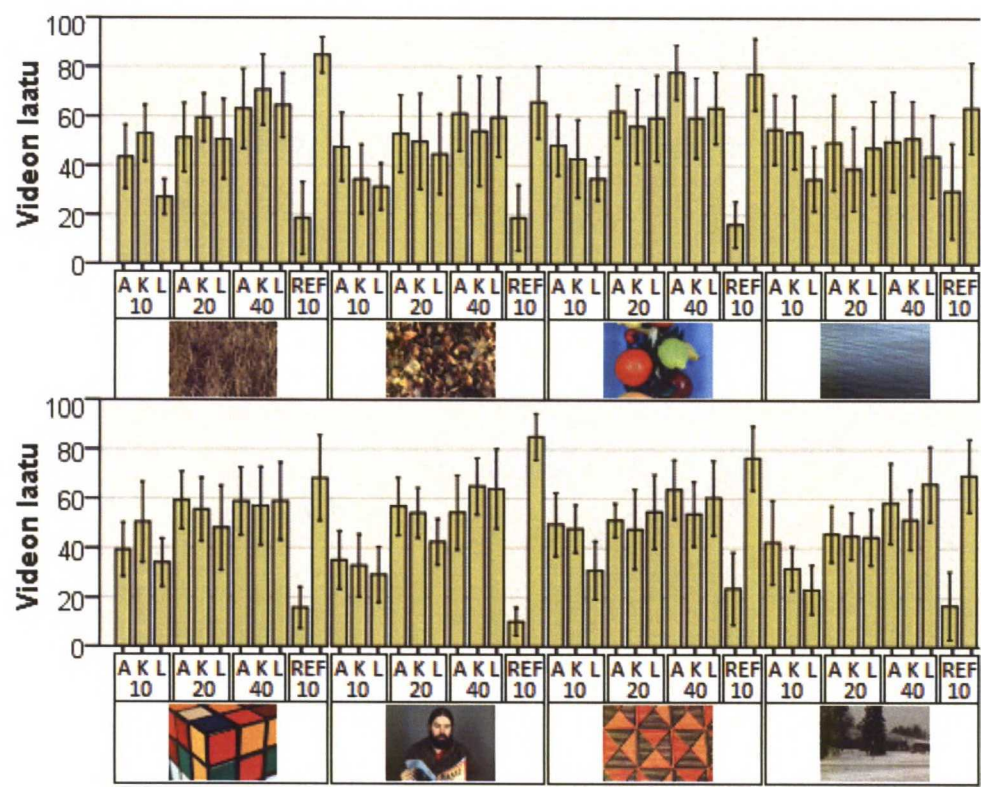
Esitestit

Taulukko A.1: Koehenkilöiden muistamat testivideot esitestien I ja II jälkeen.

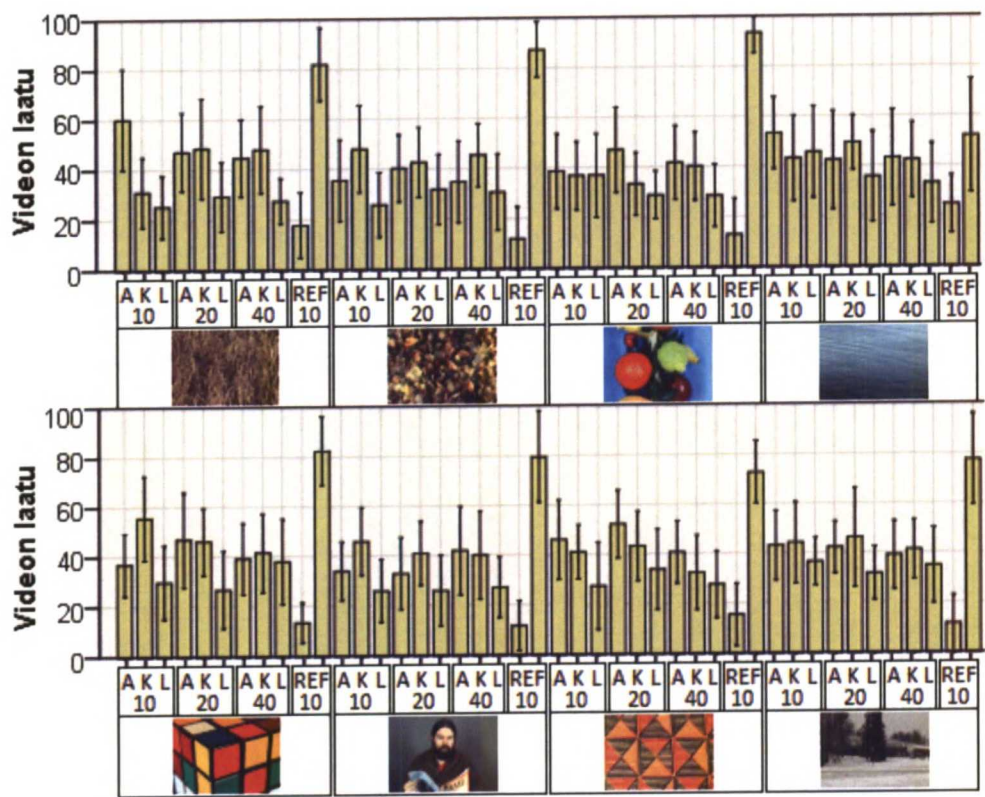
Video	Muistamisprosentti
Maisema	94%
Vesi*	92%
Lukemis	83%
Vaahtera	83%
Hyppynaru	83%
Lippu**	80%
Rubik	78%
Hanhi	78%
Auto	67%
Kynttilä	67%
Lehti	67%
Kävely	61%
Pallo	56%
Kaisla	50%
Hedelmä	44%
Pilvi***	40%
Kangas	39%
*esimerkkivideona esitestissä I	
**esimerkkivideona esitestissä II	
***ei mukana esitestissä II	

Liite B

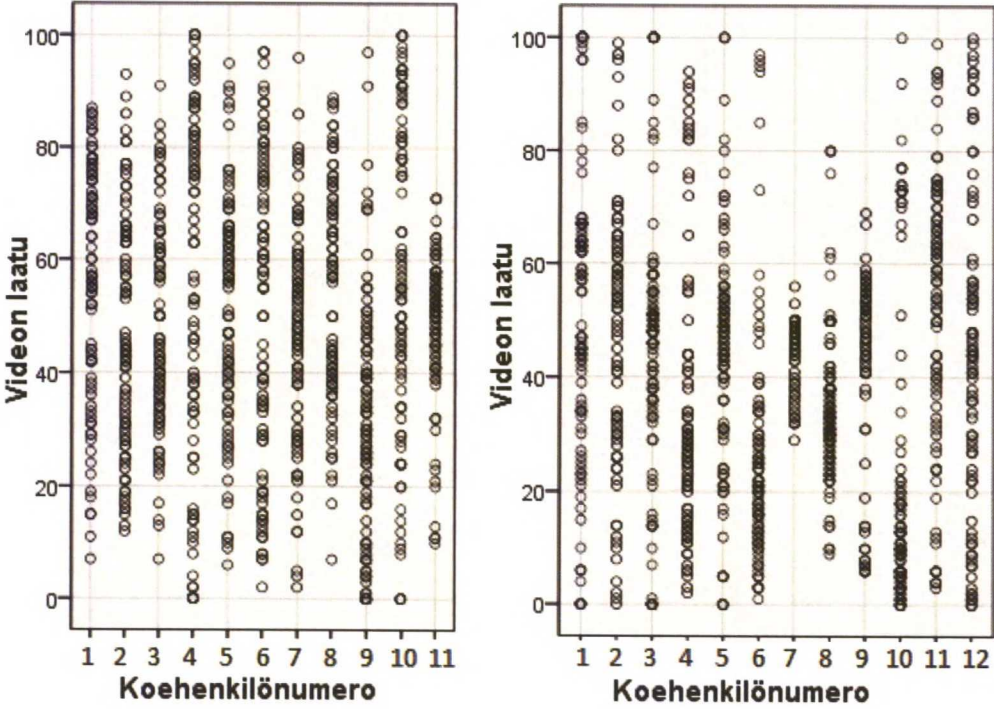
Testit I ja II



Kuva B.1: Testin I videokohtaiset tulokset. Videon pituus on joko 10,20 tai 40 sekuntia. Viisi sekuntia kestävä vääristymän sijainti on videon alussa (A), keskellä (K) tai lopussa (L).



Kuva B.2: Testin II videokohtaiset tulokset. Videon pituus on 10,20 tai 40 sekuntia. Puolet videon pituudesta kestävä vääristymän sijainti on videon alussa (A), keskellä (K) tai lopussa (L).



Kuva B.3: Testin I ja II koehenkilöiden arviojakaumat.